EFFICIENT SMALL AREA ESTIMATION IN THE PRESENCE OF MEASUREMENT ERROR IN COVARIATES

A Dissertation

by

TRIJYA SINGH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Statistics

EFFICIENT SMALL AREA ESTIMATION IN THE

PRESENCE OF MEASUREMENT ERROR IN COVARIATES

A Dissertation

by

TRIJYA SINGH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Suojin Wang
	Raymond J. Carroll
Committee Members,	Ursula Mueller-Harknett
	Gregory Berkolaiko
Head of Department,	Simon J. Sheather

August 2011

Major Subject: Statistics

ABSTRACT

Efficient Small Area Estimation in the

Presence of Measurement Error in Covariates. (August 2011)

Trijya Singh, B.A., University of Lucknow;

M.A., University of Lucknow

Co–Chairs of Advisory Committee: Dr. Suojin Wang
Dr. Raymond J. Carroll

Small area estimation is an arena that has seen rapid development in the past 50 years, due to its widespread applicability in government projects, marketing research and many other areas. However, it is often difficult to obtain error-free data for this purpose. In this dissertation, each project describes a model used for small area estimation in which the covariates are measured with error. We applied different methods of bias correction to improve the estimates of the parameter of interest in the small areas.

There is a variety of methods available for bias correction of estimates in the presence of measurement error. We applied the simulation extrapolation (SIMEX), ordinary corrected scores and Monte Carlo corrected scores methods of bias correction in the Fay-Herriot model, and investigated the performance of the bias-corrected estimators. The performance of the estimators in the presence of non-normal measurement error and of the SIMEX estimator in the presence of non-additive measurement error was also studied. For each of these situations, we presented simulation studies to observe the performance of the proposed correction procedures. In addition, we applied our proposed methodology to analyze a real life, nontrivial data set and present the re-

sults.

We showed that the Lohr-Ybarra estimator is slightly inefficient and that applying methods of bias correction like SIMEX, corrected scores or Monte Carlo corrected scores (MCCS) increases the efficiency of the small area estimates. In particular, we showed that the simulation based bias correction methods like SIMEX and MCCS provide a greater gain in efficiency. We also showed that the SIMEX method of bias correction is robust with respect to departures from normality or additivity of measurement error. We showed that the MCCS method is robust with respect to departure from normality of measurement error.

To Mom and Dad, whose patience, love and understanding never waver or fail.

ACKNOWLEDGMENTS

I would like to acknowledge and profusely thank the co-chairs of my committee, Dr. Suojin Wang and Dr. Raymond Carroll, whose support, insight and guidance have influenced me tremendously and were vital for the completion of this dissertation. I have learned a great deal from both of them, and those lessons will always remain invaluable to me during the course of my career. Without their excellent mentorship, I would not have succeeded in my goal.

I would also like to thank Dr. Simon Sheather, Dr. Michael Longnecker, Dr. Uschi Harknett, Dr. Jianxin Zhou and Dr. Gregory Berkolaiko for their unwavering support during the course of my graduate studies.

To the many fine faculty in the Department of Statistics at Texas A&M University, thank you for the fine education I have received at Texas A&M University.

Last, but definitely not the least, I would like to thank my parents and friends from the bottom of my heart for supporting me and believing in me all these years, whether they are close by or thousands of miles away.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

An overview of the dissertation is as follows. In Chapter II, we shall introduce the topics of small area estimation and measurement error models, so as to provide some background about the topics at hand. We shall very briefly mention the existing literature on measurement error in small area estimation, which is very limited.

In Chapter III, we start with an additive measurement error model for small area estimation and apply a simulation based method of bias correction called simulation extrapolation (SIMEX). We study the properties of the SIMEX estimator theoretically and via simulation. Finally, we apply this technique to a data set obtained from the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS). We also study, via simulation, the properties of the SIMEX estimator when departures from normality or additivity of measurement error occur.

In Chapter IV, we again consider an additive measurement error model, and apply the method of corrected scores, first suggested by Nakamura (1990). We study the properties of these corrected score estimators and then also apply the method of Monte Carlo corrected scores first suggested by Novick and Stefanski (2002), which is another simulation based method of bias correction. These corrected score estimators are again obtained for the same data set as in Chapter II and their performance is studied.

The journal model is *Journal of the American Statistical Association.*

It is seen from the simulation results that the ordinary corrected score estimators are as efficient or marginally better than the Lohr-Ybarra estimators. But the SIMEX and MCCS estimators perform the best, specially in the case of departure from normality or additivity of measurement error. We also show that bias corrected estimators are more efficient than those estimators in which no bias correction methods are applied even though there is presence of measurement error in covariates. When these correction methods are applied to the NHANES and NHIS body mass index data set mentioned earlier, there is a great improvement in the jackknife estimates of the mean squared errors of the estimates. Hence we achieve our goal of efficiency in small area estimation through bias correction.

CHAPTER II

REVIEW OF SMALL AREA ESTIMATION METHODS AND MEASUREMENT ERROR MODELS

In this chapter we shall provide some background on the problems of small area estimation and measurement error models and limited literature review that is of interest with respect to this dissertation before we proceed to explain how we applied bias correction methods to the problem of measurement error in covariates in this specific scenario.

2.1. Small Area Estimation

2.1.1. What is a Small Area?

Let $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2, ..., \mathcal{U}_N)$ be a finite population and let $Y_1, Y_2, ..., Y_N$ be the values of characteristic Y of our interest. Y_i could be the income of $\mathcal{U}_i$ or Y_i could be 1, if $\mathcal{U}_i$ belongs to a specified category and is zero otherwise. We are interested in estimating the population parameter's mean ($\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$), total ($Y_T = \sum_{i=1}^{N} Y_i$), median, quartiles or proportion of persons in the population who belong to a specified group. For example, we may be interested in estimating an average income or estimating the proportion of smokers in the population. In this case, $Y_i = 1$ if $\mathcal{U}_i$ smokes and zero otherwise and hence $P = \frac{1}{N}\sum_{i=1}^{N} Y_i$ = numbers of smokers/N. We may ascertain values of $\bar{Y}$, P, etc., by complete enumeration or census. But we may not have required funds, time or required funds, time or required number of experts needed for the purpose. So we conduct a large scale survey (taking n units in the sample) using appropriate designs. These designs or probability sampling schemes include simple random sampling, stratified random sampling, cluster sampling, etc.

Often it happens that we are interested to know values of the parameters for a specified subgroup or domain of the population. For example, the population may be all families in Texas. $\bar{Y}$ may be the average income of families in Texas. But we may also be interested in the average income of families in College Station, i.e., $\bar{Y}_{CS}$. Therefore, College Station households constitute a domain of our interest. There could be two types of domains:

- Planned Domain Structure: This domain is the part of population for which we plan the survey design beforehand. So domains are defined as strata in the survey design and are treated as independent. Domain sample sizes are fixed in the design itself (for example, Neymann allocation in stratified random sampling) and sampling design is also fixed for the domain (for example, a simple random sample of size n_{opt} from a given stratum. Stratification for the domain structure is an efficient option (that is, stratification of population is done to obtain a more representative sample).

- Unplanned Domain Structure: The *small areas* of our interest are, in fact, unplanned domains. These domains are specified after the sample has already been drawn using some probability sampling design. In this case, the domain structure is not a part of the sampling design as the domain is not specified at the time of planning survey design (simple random sampling, or stratified or multi-stage). Domain sizes are random variables in this case.

Some examples of small areas are:

- Geographical Area: This could be a state, province or county. It could be a district, where the state is treated as the large area of the large scale survey. It also could be a development block (here, district is taken for large scale survey). Note that each district has a number of development blocks.

- A Demographic Group: In this case, the small areas are groups of units belonging to a specified category based on age $\times$ sex $\times$ race classification in a large area (here the large scale survey is conducted at higher level), for example, black women in Texas in the age group 19-30 years. Here the large scale survey is conducted for women of this category in the U.S. and the parameter of interest may be the percentage of unemployed women.

- Geographic and Demographic Groups: An example of this kind of small area could be :"senior citizens in Texas" where the large scale survey is conducted for this group in the U.S. The parameter of interest here may be the percentage of people below poverty line in this category.

Alternative names of small areas are domains, subdomains, local area or rare domains. Some salient features of a small area are as follows:

- The large scale surveys are planned and carried out for the population (country, state or district) as a whole. The small area is specified after that. Clients ask for more than initially planned. So the sample size for the small area (that is, number of units of large scale survey already conducted falling in the small area) is a random variable. This number is usually small or even zero. No sampling frame or design or sample size is conducted for this area in the large-scale survey.

- Sample units obtained from the large scale sample for the small area may not be representative with respect to characteristic of interest.

- Large scale surveys provide reliable (unbiased, minimum variance, etc.) estimates at national level or state level (planned domain). For local-level administrative records or census data regarding units in small area (which is available) is used as covariate to produce 'small area estimates' (usually model-based estimates) to achieve given precision.

Some data sources for small area estimation might be:

- Census or complete enumeration: Information on many variables for each unit is collected at regular intervals of time. We know that population census is carried out after every 10 years. The census information is published and easily available.
- Administrative Registers: Information is collected and is maintained in office registers of different departments. These registers are well-maintained and are usually not published for public use. For example, the record of births is maintained by hospitals or tax data maintained in revenue departments.
- Large scale survey conducted in the past providing relevant information needed for small area in question.
- Satellite images: Such information is used for area-under-crop studies.

Potential data sources could be precisely divided into three broad categories: (a) data measured for the characteristics of interest in other similar areas, that is, *borrowing strength* from similar areas (b) data measured for characteristics of interest on previous occasions which includes census data (c) auxiliary information on units of small area or similar areas. One may use information from two or more sources.

Some of the issues involved in small area estimation are:

- Definition of a small area: After a large scale sample based on a given sampling design (SRS, stratified random sampling, two-stage sampling, etc) is taken for a nation or a state and data is analyzed, the client may ask for more, that is, he may desire to have estimates at lower levels like districts or counties without any further sampling. To meet the client's desire, we need to define a small area for which estimates are plausible and which also meets the client's requirement. So before defining the small area, we must understand the context.

- Identification of data sources: We must see what information is available that is relevant to our objective and information we can use for the estimation of parameters of small area.
- Method of combining of information from different data sources: We should decide what the approach will be: design-based or model-based. Both these approaches are discussed in detail later.

Bogue and Duncan (1959) remarked that small area estimation has widespread applicability, not just in government programs and funding for fair distribution of resources and fund allocation, planning and program evaluation, but also in market research by many private business houses that need small area statistics on income, consumption, habits and environment data to evaluate markets for new products, expansion or contraction of their activities. Some well-known studies involving small area estimation are:

- In the National Center for Health Statistics, U.S. National Natality Survey (1980), a national survey of $n = 9,941$ live births was taken to estimate the percentage of jaundiced live births in the country. Later, it was decided to produce estimates for the different states from this sample. Note that the survey was neither planned nor designed for state estimates. Many states had a very small number of units in the national sample so small area estimates for states were needed.

- A large scale survey of $n = 4300$ individuals was carried out for estimating drug usage (i.e., percentage of drug users) in the state of Nebraska. When this large sample data was considered for estimating the percentage of drug users in different counties of Nebraska, it was found that the Boone county has only 14 out of 4300 individuals and only 1 white female aged 25-44 years. The sample of 14 individuals cannot provide a reliable estimate of percentage of drug users in Boone nor an estimate of white women in the age group 25-44. In this study, synthetic estimation was applied. Other synthetic estimators are regression synthetic estimators and GREG estimators discussed by Rao (2003a) in detail.

An important distinction made while conducting small area estimation is between **Basic Area Level (Type A)** and **Basic Unit Level (Type B)** models. In the former, only area-specific auxiliary data, $\mathbf{X}_i = (X_{1i}, ..., X_{pi})^T$ related to some suitable function $g(Y_i)$ of the small area total Y_i for $(i = 1, 2, ..., m)$. Here, m is the number of small areas under consideration and p is the number of auxiliary variables being measured in each small area. The first attempt to use an area level model was made by Fay and Herriot (1979) to estimate the per capita income of small areas. In the second type of model, unit level auxiliary variables $\mathbf{X}_{ih} = (X_{1ih}, ..., X_{pih})^T$ are related to the unit y-values Y_{ih} through a nested error linear regression model $Y_{ih} = {X_{ih}}^T\beta + v_i + e_{ih}$ where $h = 1, ..., H$ and X_{pih} denotes the value of the p^{th} auxiliary variable observed in the i^{th} small area on the h^{th} individual. This model was first used by Battese *et al.* (1988) to model county crop areas in the U.S. Various extensions of this unit-level model have been proposed to handle binary responses, two-stage sampling within area, multivariate responses and others.

Various extensions of these models and estimation methods are discussed by Rao

(2003b). The area level Fay-Herriot model (1979) is considered in this dissertation for correction of bias due to measurement error in covariates. They proposed a model-dependent approach that uses borrowed strength from sample survey data collected in other areas to obtain an estimator for a parameter in one small area with the help of a linking model.

2.1.2. Methods of Small Area Estimation

The problem with small area estimation is that the small area sample size (number of units of large scale survey falling in the specified area) is generally too small, maybe even zero. Such a sample may give design-based unbiased (called direct estimate) but this estimate will be unstable in repeated sampling because of the low small area sample size. We need a stable estimator (minimum variance or mean squared error) so thus we have two goals: (a) to produce reliable estimates of parameters of interest (total, mean, proportion, median or quantiles) of small area for which sample size is very small (b) to assess the estimation error, that is, to produce estimates of the variance or MSE of the estimators formulated. Small area estimates are broadly of two kinds:

- Direct Estimates: The direct estimate of the parameters of small area is the unbiased (in most cases) estimator based on the sample of units of this area which are contained in the large scale survey or Current Population Survey (CPS) at any given time. This estimator may be biased if it uses auxiliary information X on the sample units which is correlated with the variable of interest Y. The direct estimators are usually design based and use survey weights and inference is done under the sampling design. These estimates are generally unbiased but have large variance because of small sample size. To enhance efficiency of direct estimate we could take another sample to increase sample size in the area but this would need more funds, time and expertise which may not be possible to financial constraints. The direct estimator uses the values of the variable of interest only for the time period of interest and units of the domain of interest only which are part of the large sample.

- Indirect Estimates: These estimates borrow strength from other domains and other time points. Three types of indirect estimators are:
  - Domain Indirect: These estimators use values of the variable of interest from units in another domain but not from another time.
  - Time Indirect: These estimators use values of the variable of interest from another time period but not from units in another domain.
  - Domain and Time Indirect: This estimator uses values of the variable of interest from another time period and from units in the other domain.

  Some examples of indirect estimators are synthetic estimators, composite estimators, EBLUP, empirical Bayes (EB) estimator, hierarchical Bayes (HB) estimator, etc.

2.2. The Fay Herriot Model

One of the most widely used area level models which is the focus of this dissertation is the Fay-Herriot model (1979) which uses the area level information concerning the characteristic of interest and the data on the auxiliary variables available from the census records or from administrative registers. If the population comprises of m mutually exclusive small areas $(\mathcal{U}_1, \mathcal{U}_2, ..., \mathcal{U}_m)$ then given the information from a current large scale national survey, we may be interested in estimating subpopulation $\mathcal{U}_i$ total

$Y_i = \sum_j Y_j$ or subpopulation mean $\bar{Y}_i = Y_i/N_i$ of small area $\mathcal{U}_i$, $i = 1, 2, .., m$ given the area level information about characteristic Y from current survey and area level auxiliary information $\underset{\sim}{X}_i^t = (X_{i1}, X_{i2}, .., X_{ip})$ on p covariates. Note that the current survey did not consider $\mathcal{U}_i$'s at the planning stage so the sample size in these small areas may be very small.

Let θ^t or $(\theta_1, \theta_2, ..., \theta_m)$ be the vector of parameters of inferential interest (for example, small area totals or means). Then the Fay-Herriot model relates the small direct design based unbiased predictor $\hat{\theta}_i$ to the area-specific auxiliary data $\underset{\sim}{X}_i^t$, through a linking model. Hereafter, we shall take $\hat{\theta}_i = y_i$, that is, y_i is the direct sample survey estimate of θ_i. The Fay-Herriot model consists of the following two models.

Level 1: Sampling model, which is,

$$y_i = \theta_i + e_i, \tag{2.1}$$

where θ_i is the unobserved small area parameter of interest and e_i's are independent sampling errors of the survey estimates with $E(e_i|\theta_i) = 0$ and $Var(\hat{\theta}_i) = \psi_i$ or σ_i^2, $i = 1, 2, ..m$. Here, ψ_i is, in fact, design-based sampling variance.

Level 2: Linking model, which is,

$$\theta_i = \underset{\sim}{X}_i^t \beta + v_i, \tag{2.2}$$

where $\underset{\sim}{X}_i^t = (X_{i1}, X_{i2}, .., X_{ip})$ is the vector of area level auxiliary information on p covariates capturing the area level effect and β is a p-vector of regression coefficients also known as fixed effects. The elements v_i's are the unobservable random effects (also called model errors) and are assumed to be independently distributed with $E(v_i) = 0$ and $Var(v_i) = \sigma_v^2$ for all i. The random effects v_i capture the additional area specific effects not explained by the area specific auxiliary variables. Thus the area level random effects v_i's capture the unstructured heterogeneity among the areas that are not explained by the sampling variances ψ_i's. This unstructured heterogeneity is exploited at the expense of an additional unknown variance component σ_v^2 to be estimated from the data. The corresponding regression model without random effect fails to capture this area-specific variability. The variance σ_v^2 is a measure of homogeneity of areas after accounting for the covariates. Thus integrating level 1 and 2 we get the Fay-Herriot model as:

$$y_i = \underset{\sim}{X}_i^t \beta + v_i + e_i, \tag{2.3}$$

for $i = 1, 2, ..., m$. The model errors v_i's and sampling errors e_i's are assumed to be independent. In many applications the normality assumption is also included. That is, $y_i|\theta_i \sim Normal(\theta_i, \psi_i)$ and $\theta_i \sim Normal(X_i^t \beta, \sigma_v^2)$. The parameters β and σ_v^2 of linking models are generally unknown and are estimated.

Remark 1: The advantage of the area-level is that it takes into account the survey design through the use of the direct estimates and related design based variance estimates. The Fay-Herriot model has the sampling model for the direct survey estimates and linking model for the small area parameters. The sampling assumes that there exists a direct survey estimator which is design unbiased for small area parameter θ_i.

Remark 2: The Fay-Herriot model is a special case of general linear mixed model:

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + Z_{n\times q}v_{q\times 1} + \epsilon_{n\times 1},$$

where $y_{n\times 1}$ is the observed vector of the response variable of interest, X and Z are matrices of explanatory variables, β is a $p\times 1$ vector of regression coefficients (fixed effects), v is a $q\times 1$ vector of unobservable random effects and ϵ is an $n\times 1$ vector of random errors. If we take $n=q=m$ where m is the number of small areas and $Z_{m\times m}=I_{m\times m}$ and consider y_i's as direct estimates of parameter of interest, we get the Fay-Herriot model.

Remark 3: The Fay-Herriot model assumes that sampling variances ψ_i's are known. Fay and Herriot (1979) have given a justification for this assumption mentioning that suitable transformations of data could be identified so as to get known sampling variances. However, in practice, smoothed estimators of the sampling variances could be obtained and then treated as known. The smoothing of variance estimates usually makes use of generalized variance function (GVF) method (Dick, 1995, Wolter, 2007). The method uses a regression model constructed for the direct estimates of sampling variances by using some external auxiliary variables and the final fitted values are obtained as the smoothing estimates. The GVF method requires an additional model for the direct estimate of the sampling variance.

Henderson (1975) showed that BLUP (best linear unbiased predictor) of θ_i is obtained as:

$$\widehat{\beta}_{GLS}=\left\{\sum_{i=1}^{m}\frac{X_iX_i^t}{\psi_i+\sigma_v^2}\right\}^{-1}\left\{\sum_{i=1}^{m}\frac{X_iy_i}{\psi_i+\sigma_v^2}\right\}, \tag{2.4}$$

and

$$\widehat{\theta}_{iBLUP}=\varphi_iy_i+(1-\varphi_i)X_i^t\hat{\beta}, \tag{2.5}$$

where $\varphi_i = \sigma_v^2/(\psi_i + \sigma_v^2)$. Note that the $\widehat{\theta}_{iBLUP}$ is the weighted average of the direct survey estimate y_i and synthetic estimator $X_i^t\hat{\beta}$ of θ_i. The weight φ_i is known as shrinkage factor. When $\varphi_i \to 0$, then $\widehat{\theta}_{iBLUP} \to X_i^t\beta$ which is the synthetic estimate of β. When $\varphi_i \to 1$ then $\widehat{\theta}_{iBLUP} \to y_i$. In practice, when the variance σ_v^2 is replaced by an estimator $\hat{\sigma}_v^2$, we get what is known as Empirical BLUP or EBLUP.

2.3. Measurement Error Methods and Models

Grace Yi (2000) remarked that measurement error has long been a concern in medical, health and epidemiological studies. It arises commonly in a variety of settings including longitudinal studies, case-control studies, survival analysis and survey sampling. Instead of observing the true value of X_i we observe an error-contaminated surrogate version or proxy variable W_i. In the regression context although both the dependent and independent variables could be measured with error but we shall concentrate on the case where the independent variable or covariate(s) are measured with error. The total measurement error could be due to some or all of these reasons: reliance on the self-reported information, use of records of suspect quality, intrinsic biological variability, laboratory analysis error, instrument error or environmental effect. Sometimes, it is impossible to measure the covariate accurately due to its nature and sometimes the covariate of interest may be difficult to measure precisely due to physical location or cost and we observe a proxy variable.

A classic example of a measurement error study is the Nevada radiation exposure study, where there were over 2000 individuals who were exposed to radiation as children due to above ground nuclear testing in the 1950's and 1960's. Lyon *et al.* (2006) commented that the response variable (Y_i) was taken to be the radiation actually observed into the thyroid. The primary radiation came from milk and vegetables. The plan of the study was to relate various thyroid disease outcomes to the radiation exposure to the thyroid. The exposure of interest, radiation to thyroid (X_i) cannot be observed exactly. A model was considered to convert the known data about the above-ground nuclear tests to the radiation actually observed into the thyroid. Dosimetry calculations were based on age at exposure, gender, residence history, X-ray history, whether the individual was breast-fed as a child and a diet questionnaire filled out by the parents focusing on the milk and vegetable consumption. The data were then input into a complex model and for each individual a point estimate of thyroid dose (X_i) and the associated standard error for the measurement error were reported. The measurement error here is due to the complex method of calculating "dose estimates of X_i's". It is justifiable here that the observed dose (W_i) equals the true dose X_i plus the measurement error U_i, which is also called classical error. The estimated doses in fact, vary around the true doses and the variability of the observed doses among the cases greater than the variability of the true doses (See Carroll *et al.*, 2006 for details).

Another example, given by Naima Shifa in her Ph.D. thesis (2009, Bowling Green State University) is as follows. Consider the relationship between the yield of corn and available nitrogen in the soil. Let the relationship between the yield (Y_i) and the level of nitrogen (X_i) be: $Y_i = \beta_0 + \beta_1 X_i + e_i$ for ($i = 1, 2, ..., n$). The coefficient β_1 is the amount by which the yield will increase or decrease when the soil nitrogen increases or decreases by one unit. To estimate the available soil nitrogen, we need to sample the soil of experimental plot and perform a laboratory experiment. Thus, because of sampling and laboratory experiment, measurement error is introduced and we observe W_i instead of X_i.

For understanding the effect of measurement error and the statistical methods for analyzing data with error, we explain the following models based on the error structure:

- *Classical measurement error model*: Here, we measure contaminated W_i, instead of the true value X_i and $W_i = X_i + U_i$, where U_i is an additive error. We have reason to believe that W_i depends on X_i.
    - U_i has mean zero and is independent of X_i and response variable Y_i and U_i has constant variance σ_u^2 .
    - If U_i is independent of all variables in the model (X_i, Y_i, Z_i, etc.) then it is non-differential and W is a surrogate.
    - W_i is thought to fluctuate around the true covariate X_i and the variability in W_i is more than the variability in the X_i.
    - $E(W_i|X_i) = X_i$, that is W_i is an unbiased estimator of X_i. Note that X_i could be a random variable that is why we consider its conditional expectation. More generally, we may have $W_i = \nu_0 + \nu_1 X_i + U_i$, which is a case where unbiasedness of W_i is not tenable.

  M.S. Hossain (Ph.D. thesis, University of British Columbia, 2007) quotes a study on cholesterol level and coronary heart disease where $W_i = \nu_0 + \nu_1 X_i + U_i$ is a more appropriate model. In this study, the risk of coronary heart disease (CHD) is treated as a function of blood cholestrol level. The main interest is the effect of low density lipoprotein cholesterol (LPL) on CHD. But the X_i (LDL) is not measured as its measurement error is more costly than that of the total cholesterol. The CHD status is taken as the dependent variable Y , LDL/100 as independent variable X but $TC/100 = W$ is measured. Note that

LDL level constitutes a part of TC level $TC = LDL + K$ and $E(K) \neq 0$. Thus the assumption of unbiased measurement error does not hold in this case.

- *Berkson measurement error model*: If the observed values (which are actual measurements) W_i's are fixed in repeated sampling and the true values X_i vary around W_i, then the appropriate model when we believe that X_i depends on W_i is $X_i = W_i + U_i$ where U_i has mean zero and is independent of W_i. Here, the realization of W_i comes before X_i. Such a model usually arises in the experimental situation with controlled mechanism where the observed variable W_i is controlled, so this model is also known as the controlled variable model. The fixed values (observed W_i's) are known but the true X_i's are unknown.
  - $E(U_i|X_i) = 0$ regardless of the value of W_i as W_i could also be random. Thus $E(X_i|W_i) = W_i$. Hence W_i is called an unbiased Berkson predictor of X_i.
  - In this model, X_i's (true values) are more variable than W_i.

In the Berkson model, we have $E(X_i|W_i) = W_i$. However, the variability is different. We may also have a situation where $X_i = \nu_0 + \nu_1 X_i + W_i$. Buzas *et al.* (2005), in their review article of measurement error mention an example where in an experimental design for curing a material in kiln at a specified temperature W_i is set by the thermostat. Although the thermostat is set at W_i, the actual temperature in the kiln, X_i, often varies randomly from W_i due to less than perfect thermostat control. For a properly calibrated thermostat, a reasonable assumption is $E(X_i|W_i) = W_i$, which is a salient feature of Berkson measurement error.

2.3.1. Estimation Methods

The estimation methods available for measurement error models can be broadly classified into two categories.

Functional Estimation Methods: Let us suppose that the classical measurement error model described previously exists. Here, we assume that the response variable Y is measured without error. The variable X that is measured with error is not considered to be random but fixed, so the $X_1, X_2, ..., X_n$ form a sequence of unknown fixed constants and hence so are the parameters pertaining to the observations $W_1, W_2, ..., W_n$, also called incidental parameters. Under these assumptions, the method used for estimation of parameters is called a functional estimation method and such a model is known as a functional model. Thus, the unobserved covariates are modeled as unknown random constants (hence parameters). In this case the number of parameters grows with the sample size and thus poses many problems like model identifiability, inconsistency of m.l.e.'s of parameters as consistency of parameters remain fixed as the sample size increases and due to the non-existence of a global maxima for m.l.e.'s. The solution of estimating equations may be local minima or saddlepoint so in each case we have to verify the conditions for valid m.l.e.'s. For the valid parameter estimates, we may have to impose certain conditions on the parameters or on a function of the parameters (Cheng and Van Ness, 1999).

An alternative variant of functional model (Carroll *et al.*, 2006) is not to use the distribution of X_i's even if such a distribution exists. We circumvent problems posed above due to presence of incidental parameters by directly estimating regression parameter β using iid observations $(W_i,\, Y_i)$, $i = 1, 2, .., n$. In this approach, X_i's may be either fixed or random but in the latter case no or minimal assumptions are made about the distributions of X_i's. The advantage of this approach is that it leads to the estimation procedures which are immune against the possible misspecification of the distribution and are also valid when X_i's are non-stochastic.

The methods of least squares or maximum likelihood usually used to estimate regression parameter β provide biased or incorrect estimates when the covariates are measured with error. In fact, bias is introduced in the estimation equation itself, and that in turn translates into the bias in the estimates of β. In order to reduce bias under functional modeling, we use methods like, (i) corrected scores and Monte Carlo corrected scores, (ii) simulation extrapolation (SIMEX), (iii) conditional scores function methods (Stefanski and Carroll, 1990), based on the theory of sufficient statistics on which we can condition to eliminate nuisance parameters.

Structural Estimation Methods: Under this approach we assume that the latent variables X_i's are iid random variables. Here (X_i, Y_i) vary jointly in repeated sampling. In this case the distribution of X_i is specified by a pdf $g(X_i/\tau)$ and are independent of errors U_i's or equation errors. Note that the distribution $g(X_i/\tau)$ contains an unknown (nuisance) parameter τ. This parameter can be estimated from the data W_i's alone without recourse to the regression model. For example, if $X_i \sim Normal(\mu_x, \sigma_x^2)$ then the parameters μ_x and σ_x^2 can be estimated by $\bar{W}$ and $S_w^2 - \sigma_u^2$ respectively, where σ_u^2 is the known error variance. Then replacing τ by its consistent estimator $\hat{\tau}$ does not alter the consistency property of β, though it does have an effect on the asymptotic variance. Under structural modeling, the methods used to tackle the bias problem include the maximum likelihood estimator, the quasi-score estimator, and the regression calibration estimator.

An example of a functional model is as follows. Consider the relationship between aquatic species diversity Y and acid neutralizing capacity X, given measurements (Y_i, W_i) where W_i is observed instead of X_i for n lakes. If the only lakes of interest are those represented in the sample, then it is appropriate to model X_i as known constants and the model used is the functional estimation model. On the other hand, if the lakes represented in the data are a random sample from a large population of lakes, then it is appropriate to model X_i's as independently and identically distributed random variables and hence we have a structural estimation model.

2.3.2. Types of Measurement Error

Non-Differential Errors: Let Y be the response variable which is measured without error and let X be the true covariate measured with error. Instead of measuring X, we observe a contaminated version W, which is said to be non-differential if:

$$f(Y|X, W) = f(Y|X).$$

Moreover, if we have two covariates X and Z, where X is measured with error but Z is not, then the measurement error in W is said to be non-differential if:

$$f(Y|X, W, Z) = f(Y|Z, X).$$

Both the above equations imply that in the probability sense, W contains no information about Y (or in predicting Y) in addition to whatever information is contained in X (in the first case) or X and Z (in the second case). In other words, W is conditionally independent of Y given the true covariates X and Z. Thus, non-differential measurement error means that Y depends on true covariates X and Z, but not on observed W given X and Z. This also means that given the true covariates X and

Z, Y and W are independent.

An example of differential measurement error, given by Buzas *et al.* (2005) is as follows: when covariate X is defined as the average value of time varying risk factor like blood pressure, cholesterol level, etc., or the average value of spatially varying exposure like lead exposure. The observed W is a measurement at a single point of time and space. In such cases it might be convincingly argued that the single measurement contributes little or no information in addition to that contained in the long run average or the spatial average.

Implication: With non-differential measurement error, it is possible to estimate parameters in the model relating the response to the true predictor using the measured covariate only with minimal additional information on the error distribution and it is not necessary to observe the true predictor X.

Differential Errors: The error in W is called differential if:

$$f(Y|X, W) \neq f(Y|X).$$

Moreover, if we have two covariates X and Z, where X is measured with error but Z is not, then the measurement error in W is said to be differential if:

$$f(Y|X, W, Z) \neq f(Y|Z, X).$$

Differential measurement error occurs when the response Y is obtained first and then in subsequent follow-ups, we obtain the covariates. In nutrition studies, Jones *et al.* (1987) mentioned that this ordering of measurement typically causes differential error. For instance, here the true predictor (covariate) would be long-term diet before diagnosis but the nature of case-control studies is that reported diet is obtainable only after diagnosis. A woman who develops breast cancer may well change her diet, so the reported diet is measured after diagnosis (which is different from long-term average diet) will be correlated with the disease status Y_i.

Implication: Under a differential error mechanism, the true covariate (X_i's) are not sufficient to explain response variable Y_i. The information carried by the observed measurements W_i cannot be ignored. So here, it is necessary to have a validation sub-sample in which both the measured values and the true values are recorded.

Remark: W is called a surrogate variable for X if $f(Y|X,W) = f(Y|X)$ or $f(Y|X,W,Z)$ $= f(Y|Z,X)$. Surrogate status can depend on the particular model being fit to the data. For example, consider the model where $Z = (Z_1, Z_2)$. It is possible to have:

$$f(Y|Z_1, Z_2, X, W) = f(Y|Z_1, Z_2, X),$$

but

$$f(Y|Z_1, X, W,) \neq f(Y|Z_1, X).$$

Thus W is a surrogate in the full model that includes Z_1 and Z_2 but not in the reduced model containing just Z_1 only. In other words W is a surrogate or not depends on other variables in the model.

The effect of measurement error has been referred to by Carroll *et al.* (2006) as the triple whammy since it majorly has three effects. It causes bias in parameter estimation for statistical models, leads to a loss of power for detecting relationship among variables and it masks the features of the data, making graphical model analysis difficult. The bias caused in the slope estimate due to measurement error in the direction of zero is commonly referred to as attenuation or attenuation to the null. The effects of measurement error can range from simple attenuation to situations where real effects are hidden, observed data exhibit relationships that are not present in the error free data and even signs of estimated coefficients are reversed.

2.4. The Lohr-Ybarra Model

Lohr and Ybarra (2008) for the first time in the literature associated with small area estimation, incorporated measurement error into Fay-Herriot's model, i.e., they assumed that the auxiliary information provides an estimator $\widehat{X_i}$ of the p-vector X_i of population characteristics. The estimator $\widehat{X_i}$ has a mean squared error $MSE(\widehat{X_i}|X_i) = C_i$ under the sampling design. In our model we consider $\widehat{X_i} = X_i + U_i$, where U_i is the measurement error for the auxiliary information in the i^{th} small area and $U_i \sim N(0, C_i)$. Their model is:

$$y_i = \widehat{X}_i^T \beta + r_i(\widehat{X}_i, X_i) + e_i, \tag{2.6}$$

where $r_i(\widehat{X}_i, X_i) = v_i + (X_i - \widehat{X}_i)^T \beta$, $v_i \sim (0, \sigma_v^2)$ and $e_i \sim (0, \psi_i)$

Remark: This model is obtained as follows. The Fay-Herriot model is $y_i = X_i^T \beta + v_i + e_i$. Hence, keeping in mind that $\widehat{X}_i = X_i + U_i$ and $r_i(\widehat{X}_i, X_i) = v_i + (X_i - \widehat{X}_i)^T \beta$ we have:

$$\begin{aligned} y_i &= X_i^T \beta + v_i + e_i \\ &= (\widehat{X}_i - U_i)^T \beta + v_i + e_i \\ &= \widehat{X}_i^T \beta - (\widehat{X}_i - X_i)^T \beta + v_i + e_i \\ &= \widehat{X}_i^T \beta + r_i(\widehat{X}_i, X_i) + e_i. \end{aligned}$$

They assumed that $\widehat{X}_i$ as an estimator of X_i exists for each area i. If some components of X_i are not measured, they suggested using imputed values or estimates. Of course, if a component $\widehat{X}_{ih}$ is known exactly, $\widehat{X}_{ih} = X_{ih}$ is taken. They also assumed that v_i is independent of both $\widehat{X}_i$ and e_i and that random variables in different small areas are independent. For simplification they assumed that $\widehat{X}_i$ and y_i are independent for each area i.

Under the model suggested by Ybarra and Lohr, they proposed an estimator given by:

$$\widetilde{Y}_{iME} = \gamma_i y_i + (1-\gamma_i)\widehat{X}_i^T\beta, \tag{2.7}$$

where,

$$\gamma_i = \frac{\sigma_v^2 + \beta^T C_i \beta}{\sigma_v^2 + \beta^T C_i \beta + \psi_i} = \frac{MSE(r_i)}{MSE(r_i) + \psi_i},$$

and $r_i = r_i(\widehat{X}_i, X_i) = v_i + (X_i - \widehat{X}_i)^T\beta$

The predictor they proposed under this model is a convex, linear combination, i.e., a weighted average of the direct estimator, y_i, and the predicted value $\widehat{X_i}^T\beta$. The Lohr-Ybarra estimator reduces to the Fay-Herriot estimator when the auxiliary information is measured with no error. They showed that this estimator has minimum mean squared error amongst all linear combinations of y_i and $\widehat{X}_i^T\beta$ of the form $a_i y_i + (1 - a_i)\widehat{X}_i^T\beta$.

Lohr and Ybarra used modified least squares to estimate the parameters (Cheng and Van Ness, 1999). They suggested that a set of finite weights bounded away from 0, $(w_1, ..., w_m)$, be considered (for $i = 1, ..., m$) and solved the following equation to obtain the estimates of the regression parameters:

$$\sum_{i=1}^{m} w_i(X_i^T - C_i)\beta = \sum_{i=1}^{m} w_i\widehat{X_i}y_i. \tag{2.8}$$

So the weighted least squares estimator of β they used was,

$$\widehat{\beta}_w = \left\{ \sum_{i=1}^{m} w_i ({X_i}^T - C_i) \right\}^{-1} \widehat{X_i} y_i. \tag{2.9}$$

The Fay-Herriot model uses weights $\widehat{w}_i = (\widehat{\sigma}_v^2 + \psi_i)^{-1}$ and Lohr-Ybarra use weights $w_i = (\sigma_v^2 + \psi_i + \beta^T C_i \beta)^{-1}$. They prove that the estimated weights $\widehat{w}_i$ are consistent. They initially set the weights to be 1 and then estimate β and σ_v^2, using Equation (2.9) and:

$$\widehat{\sigma}_v^2 = (m-p)^{-1} \sum_{i=1}^{m} \{ (y_i - \widehat{X_i}^T \widehat{\beta}_w)^2 - \psi_i - \widehat{\beta_w}^T C_i \widehat{\beta_w} \}. \tag{2.10}$$

Further, they proved that the weighted least squares estimator of β is consistent and has an asymptotic normal distribution and that the estimator of σ_v^2 is also consistent under certain regularity conditions. They used the estimated weights $\widehat{\gamma}_i$ (obtained after substituting the estimates of β and σ_v^2 in γ_i) to propose the composite measurement error estimator, $\widehat{Y}_{iME}$ of Y_i.

The problem with the Lohr-Ybarra estimator is that it is not efficient so that their estimated mean square errors are sometimes greater than even the direct estimator y_i. In this dissertation, we apply bias correction techniques to successfully tackle this problem of inefficiency of small area estimates in the presence of measurement error in covariates.

2.5. Corrected Scores Approach for a Special GLMM Model for Small Area Estimation

Zhong *et al.* (2002) derived corrected scores estimators for a general linear mixed model. In their model, however, the v_i's, the random effects of the model were not assumed to be independent. We show here how their model could be modified with respect to error structure and adapted for small area estimation. Let there be m small areas in the population from which a large scale sample has been drawn. Let the parameters of interest in these areas be $\hat{\theta}_i$, $i = 1, 2, .., m$ for which the direct survey estimates y_i's based on the sampling design are available. That is, $\hat{\theta}_i$ for $i = 1, 2, .., m$.. Consider the general linear mixed model,

$$y_i = \underset{\sim}{x}_i^t + v_i + \epsilon_i, \tag{2.11}$$

where $\underset{\sim}{x}_i^t = (x_{i1}, x_{i2}, ..., x_{ip})$ is a vector of area level values of p covariates for a small area i which are available from census or administrative records. β is a $p \times 1$ vector of regression coefficients (fixed effects). The variable v_i is the random area-specific effect and ϵ_i. v_i's are independently distributed as normal with mean zero and variance $\lambda_i \sigma_v^2$ and ϵ_i's are independently and identically distributed as normal with mean zero and variance ψ_i. Further, v_i and ϵ_i are independent. Since both types of errors correspond to the small area, the assumption seems to be reasonable. The other error structures, we shall consider in the next sections. For m areas taken together model (2.11) could be written as:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{IV} + \epsilon \tag{2.12}$$

where $\mathbf{Y}$ is an $m \times 1$ vector, $\mathbf{X}$ is an $m \times p$ matrix of known coordinates, β is a $p \times 1$ vector of fixed effects and $\mathbf{I}$ is the identity matrix of order $m \times m$. $\mathbf{V}$ is an $m \times 1$ vector of area-specific random effects and ϵ is an $m \times 1$ vector of sampling errors. $\mathbf{V} \sim Normal_m(\underset{\sim}{0}, \sigma_v^2 \Sigma)$ and $\epsilon \sim Normal_m(\underset{\sim}{0}, G)$ where $G = Diag(\sigma_1^2, ..., \sigma_m^2)$ and $\mathbf{V}$ and ϵ are independent.

Let us assume that covariates $\mathbf{X}_1, \mathbf{X}_2, .., \mathbf{X}_p$ are measured with error, so instead of measuring X_i we observe $W_i = X_i + U_i$, for $i = 1, 2, ..., m$. The error is assumed to non-differential, classical additive error. Thus,

$$\mathbf{W}_{m\times p} = \begin{pmatrix} \underset{\sim}{W}_1^t \\ \underset{\sim}{W}_2^t \\ : \\ : \\ \underset{\sim}{W}_m^t \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}_1^t \\ \underset{\sim}{X}_2^t \\ : \\ : \\ \underset{\sim}{X}_m^t \end{pmatrix} + \begin{pmatrix} \underset{\sim}{U}_1^t \\ \underset{\sim}{U}_2^t \\ : \\ : \\ \underset{\sim}{U}_m^t \end{pmatrix} = \mathbf{X} + \mathbf{U} \tag{2.13}$$

$\mathbf{U}$ is distributed as $Normal\{\mathbf{0}, I\otimes\Lambda\}$, $\underset{\sim}{U}_i$ is a p-vector with distribution $Normal\{0, \Lambda_{p\times p}\}$. $\underset{\sim}{U}_i$'s are independent of v_i's and ϵ_i's.

We assume that σ_v^2, Σ and Λ are known, but can be replaced by their consistent estimators in practical applications. Given the data, the likelihood function for the model specified by Equations (2.12) and (2.13) can be expressed as,

$$\begin{aligned} L(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y}) &= f(\underset{\sim}{\mathbf{V}})f(y/\mathbf{V},\mathbf{X},\mathbf{Y}) \\ &= K^{-1}(\sigma_v^2,\Sigma)exp\left\{\frac{-1}{2\sigma_v^2}\mathbf{V}^T\Sigma^{-1}\mathbf{V}\right\}. \\ &\quad \times exp\left\{\frac{-1}{2\sigma_v^2}[\mathbf{Y}-\mathbf{X}\beta-\mathbf{IV}]^T[\mathbf{Y}-\mathbf{X}\beta-\mathbf{IV}]\right\} \end{aligned} \tag{2.14}$$

where,

$$K^{-1}(\sigma_v^2,\Sigma) = (2\pi\sigma_v^2)^m|\Sigma|^{1/2} \tag{2.15}$$

The log-likelihood is given by:

$$\begin{aligned} l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y}) = -(m)ln(2\pi\sigma_v^2) - \frac{1}{2}ln|\Sigma| - \frac{-1}{2\sigma_v^2}[\mathbf{Y}-\mathbf{X}\beta-\mathbf{IV}]^T[\mathbf{Y}-\mathbf{X}\beta-\mathbf{IV}] \\ - \frac{1}{2\sigma_v^2}\mathbf{V}^T\Sigma^{-1}\mathbf{V}. \end{aligned} \tag{2.16}$$

Let

$$g(\sigma_v^2,\Sigma) = -m \times ln(2\pi\sigma_v^2) - \frac{1}{2}ln|\Sigma|. \tag{2.17}$$

After extensive algebra, the log-likelihood function in Equation (2.16) can be expressed as,

$$\begin{aligned} l(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}) = g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2}\left\{(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) - 2\mathbf{V}^T(\mathbf{Y} - \mathbf{X}\beta)\right\} \\ - \frac{1}{2\sigma_v^2}[\mathbf{V}^T(I + \Sigma^{-1})\mathbf{V}] \end{aligned} \tag{2.18}$$

Henderson (1975) and Robinson (1991) proposed the best linear unbiased prediction approach to estimate the fixed effects β and the random effects $\mathbf{V}$ by solving the equations:

$$\frac{\partial}{\partial \mathbf{V}} l(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}) = 0, \tag{2.19}$$

and

$$\frac{\partial}{\partial \beta} l(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}) = 0. \tag{2.20}$$

We shall use the following standard formula for partial derivatives in our work (See, 'The Matrix Cookbook'):

$$\frac{\partial}{\partial \underset{\sim}{X}}(B\underset{\sim}{X} + b)^T C(D\underset{\sim}{X} + d) = B^t C(D\underset{\sim}{X} + d) + D^t C^t(B\underset{\sim}{X} + b) \tag{2.21}$$

Using Equation (2.21) we obtain,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{V}} l(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}) &= -\frac{1}{2\sigma_v^2}[-2(\mathbf{Y} - \mathbf{X}\beta - \mathbf{IV})] - \frac{\Sigma^{-1}\mathbf{V}}{\sigma_v^2} \\ &= \frac{1}{\sigma_v^2}(\mathbf{Y} - \mathbf{X}\beta) - (I + \Sigma^{-1})\frac{V}{\sigma_v^2}. \end{aligned} \tag{2.22}$$

Equating the partial derivative in Equation (2.22) to zero, we obtain the estimates of $\underset{\sim}{V}$ as,

$$\hat{V} = (I + \Sigma^{-1})^{-1}(\mathbf{Y} - \mathbf{X}\beta). \tag{2.23}$$

Substituting this value of $\hat{V}$ in Equation (2.18) we obtain,

$$\begin{aligned} l(\beta;\hat{V}(\beta),\mathbf{X},\ \mathbf{Y}) = g(\sigma_v^2,\Sigma) &- \frac{1}{2\sigma_v^2}(\mathbf{Y}-\mathbf{X}\beta)^T(\mathbf{Y}-\mathbf{X}\beta) \\ &+ \frac{1}{\sigma_v^2}[(I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{X}\beta)]^T(\mathbf{Y}-\mathbf{X}\beta) \\ &+ \frac{1}{2\sigma_v^2}\Big\{[(I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{X}\beta)]^T \\ &(I+\Sigma^{-1})[(I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{X}\beta)]\Big\}. \end{aligned} \tag{2.24}$$

After extensive algebra Equation (2.24) simplifies to,

$$\begin{aligned} l(\beta;\hat{V}(\beta),\mathbf{X},\ \mathbf{Y}) &= g(\sigma_v^2,\Sigma) \\ &- \frac{1}{2\sigma_v^2}(\mathbf{Y}-\mathbf{X}\beta)^T[I+(I-\Sigma^{-1})^{-1}](\mathbf{Y}-\mathbf{X}\beta). \end{aligned} \tag{2.25}$$

We know that if we take $R = I_{m\times m} + \Sigma_{m\times m}$ then,

$$R^{-1} = I + (I-\Sigma^{-1})^{-1}, \tag{2.26}$$

using result by C.R. Rao (1973, page 33). Substituting from Equation (2.26) in Equation (2.25), we obtain:

$$\begin{aligned} l(\beta;\mathbf{X},\ \mathbf{Y}) &= g(\sigma_v^2,\Sigma) - \frac{1}{2\sigma_v^2}(\mathbf{Y}-\mathbf{X}\beta)^T R^{-1}(\mathbf{Y}-\mathbf{X}\beta) \\ &= g(\sigma_v^2,\Sigma) - \frac{1}{2\sigma_v^2}\Big\{\mathbf{Y}^t\mathbf{R}^{-1}\mathbf{Y} - 2\beta^t\mathbf{X}^t\mathbf{R}^{-1}\mathbf{Y} + \beta^t\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X}^t\beta\Big\}. \end{aligned} \tag{2.27}$$

For known Σ and σ_v^2 the expression in Equation (2.27) depends on only the unknown parameter β. When the regressor variables $\mathbf{X}$ are measured with error as mentioned in Equation (2.13), the correlated structure arises from the random effects. Therefore, the global expectations of the score functions $\frac{\partial}{\partial \mathbf{V}}l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})$ obtained from Equation (2.16) and, $\frac{\partial}{\partial \beta}l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})$ obtained from (2.17) after replacing $\underset{\sim}{\mathbf{W}}$ shall generally not be equal to zero. Therefore estimators obtained from the score functions $S_1(\beta;\mathbf{V},\ \mathbf{W},\ \mathbf{Y}) = \frac{\partial}{\partial \mathbf{V}}l(\beta;\mathbf{V},\ \mathbf{W},\ \mathbf{Y})\Big|_{X=W}$ and $S_2(\beta;\mathbf{X},\ \mathbf{Y}) = \frac{\partial}{\partial \beta}l(\beta;\mathbf{X},\ \mathbf{Y})\Big|_{X=W}$

are not consistent in general. The estimates will be biased due to the presence of measurement error. So we shall apply the corrected scores method suggested by Stefanski and Nakamura to this model. Some other important contributions in this area are due to Carroll and Stefanski (1987a, 1987b), Gimenez *et al.* (1987) and Zhong *et al.* (2002).

Let E^* and Var^* be the respective conditional means and variances with respect to $\mathbf{W}$ given $\mathbf{V}$ and $\mathbf{Y}$, then corrected log- likelihoods $l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ and $l^*(\beta; \mathbf{W},\ \mathbf{Y})$ should satisfy the following conditions.

$$E^* \left\{ \frac{\partial}{\partial \underline{\mathbf{V}}} l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) \right\} = \frac{\partial}{\partial \underline{\mathbf{V}}} l(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}), \tag{2.28}$$

and

$$E^* \left\{ \frac{\partial}{\partial \beta} l^*(\beta; \mathbf{W},\ \mathbf{Y}) \right\} = \frac{\partial}{\partial \beta} l(\beta; \mathbf{X},\ \mathbf{Y}). \tag{2.29}$$

Now, using independence of $\underline{\mathbf{V}}$, $\underline{\mathbf{U}}$and ϵ we can easily see that $Cov(\mathbf{Y},\ \mathbf{W}) = Cov(y, \underline{\mathbf{V}})$ $= Cov(W, \underline{\mathbf{V}}) = 0$. Moreover, the marginal distributions of the rows of $\mathbf{W}$, $\mathbf{Y}$ and $\mathbf{V}$ is normal. Hence for convenience we write:

$$\begin{aligned} \mathbf{W} &\sim N\{\mathbf{X}, I \otimes \Lambda\} \\ \mathbf{V} &\sim N\{\underline{0}, \sigma_v^2 \Sigma\} \\ \mathbf{Y} &\sim N\{\underline{0}, \mathbf{G} + \sigma_v^2 \Sigma\} \end{aligned}$$

Therefore, the joint distribution of $(\mathbf{W},\ \mathbf{Y},\ \mathbf{V})$ is multivariate normal with mean vector $(\mathbf{X}, \mathbf{X}\beta, \mathbf{0})$ and variance covariance matrix given by ;

$$\Sigma^* = \begin{pmatrix} I \otimes \Lambda & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} + \sigma_v^2 \Sigma & \sigma_v^2 \Sigma \\ \mathbf{0} & \sigma_v^2 \Sigma & \sigma_v^2 \Sigma \end{pmatrix}.$$

It is easy to verify that the conditional distribution of the rows of $\mathbf{W}$ is multivariate normal $N(\mathbf{X}, I \otimes \Lambda)$. Now, replacing $\mathbf{X}$ by $\mathbf{W}$ in Equation (2.27) and taking conditional expectation with respect to $\mathbf{W}$ given $\mathbf{V}$ and $\mathbf{Y}$ we obtain:

$$\begin{aligned} E^* \left\{ l(\beta; \mathbf{W}, \mathbf{Y}) \right\} = g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2} \left\{ \mathbf{Y}^t \mathbf{R}^{-1} \mathbf{Y} - 2\beta^t E^*(\mathbf{W}^t) \mathbf{R}^{-1} \mathbf{Y} \right\} \\ + \beta^t E^*(\mathbf{W}^t \mathbf{R}^{-1} \mathbf{W}) \beta. \end{aligned} \tag{2.30}$$

The conditional distribution of $\mathbf{W}$ is multivariate normal $N(\mathbf{X}, I_{m \times m} \otimes \Lambda)$. Using the result of expectation of a quadratic form, we can easily show that:

$$E^*(\mathbf{W}^t \mathbf{R}^{-1} \mathbf{W}) = \mathbf{X}^t \mathbf{R}^{-1} \mathbf{X} + tr(R^{-1}) \Lambda. \tag{2.31}$$

Thus, substituting from Equation (2.31) in Equation (2.30) we get:

$$\begin{aligned} E^* \left\{ l(\beta; \mathbf{W}, \mathbf{Y}) \right\} &= g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2} [\mathbf{Y}^t \mathbf{R}^{-1} \mathbf{Y} - 2\beta^t \{\mathbf{X}^t \mathbf{R}^{-1} \mathbf{Y}\} \\ &\quad + \beta^t \{\mathbf{X}^t \mathbf{R}^{-1} \mathbf{X} + tr(R^{-1}) \Lambda\} \beta] \\ &= g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2} [\mathbf{Y}^t \mathbf{R}^{-1} \mathbf{Y} - 2\beta^t \{\mathbf{X}^t \mathbf{R}^{-1} \mathbf{Y} + \beta^t \mathbf{X}^t \mathbf{R}^{-1} \mathbf{X} \beta\}] \\ &\quad - \frac{1}{2\sigma_v^2} \beta^t tr(R^{-1}) \beta. \end{aligned} \tag{2.32}$$

Therefore the corrected log-likelihood $l^*(\beta; \mathbf{W}, \mathbf{Y})$ is given by,

$$\begin{aligned} l^*(\beta; \mathbf{W}, \mathbf{Y}) = g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2} [\mathbf{Y}^t \mathbf{R}^{-1} \mathbf{Y} - 2\beta^t \mathbf{W}^t \mathbf{R}^{-1} \mathbf{Y} + \beta^t \mathbf{W}^t \mathbf{R}^{-1} \mathbf{W} \beta] \\ + \frac{1}{2\sigma_v^2} tr(R^{-1}) \beta^t \Lambda \beta. \end{aligned} \tag{2.33}$$

Differentiating Equation (2.33) with partially respect to β we obtain:

$$\begin{aligned}
\frac{\partial}{\partial\beta} l^*(\beta;\mathbf{W},\,\mathbf{Y}) &= -\frac{1}{\sigma_v^2}[-\mathbf{W}^t\mathbf{R}^{-1}\mathbf{Y} + \mathbf{W}^t\mathbf{R}^{-1}\mathbf{W}\mathbf{X}\beta] \\
&\quad + \frac{1}{\sigma_v^2} tr(R^{-1})\Lambda\beta \\
&= \frac{\mathbf{W}^t R^{-1}}{\sigma_v^2}(\mathbf{Y} - \mathbf{W}\beta) + \frac{1}{\sigma_v^2} tr(R^{-1})\Lambda\beta \\
&= \Psi^*(\beta;\mathbf{W},\,\mathbf{Y}),
\end{aligned} \tag{2.34}$$

which is the corrected score function for β. The corrected observed information for β is given by,

$$I^*(\beta;\mathbf{W},\,\mathbf{Y}) = -\frac{\partial^2}{\partial\beta\partial\beta^t} l^*(\beta;\mathbf{W},\,\mathbf{Y}) = \frac{\mathbf{W}^t V^{-1}\mathbf{W}}{\sigma_v^2} - \frac{tr(R^{-1})\Lambda}{\sigma_v^2}, \tag{2.35}$$

and the corrected Fischer's information is given by:

$$\begin{aligned}
I(\beta;\mathbf{X}) &= E^+ E^*\{I^*(\mathbf{W},\,\mathbf{Y})\} \\
&= E^+\left[E^*\left(\frac{\mathbf{W}^t R^{-1}\mathbf{W}}{\sigma_v^2} - \frac{tr(R^{-1})\Lambda}{\sigma_v^2}\right)\right] \\
&= E^+\left[\frac{\mathbf{X}^t R^{-1}\mathbf{X}}{\sigma_v^2} + \frac{tr(R^{-1})\Lambda}{\sigma_v^2} - \frac{tr(R^{-1})\Lambda}{\sigma_v^2}\right] \\
&= \frac{\mathbf{X}^t R^{-1}\mathbf{X}}{\sigma_v^2}.
\end{aligned} \tag{2.36}$$

Now we shall verify the condition at Equation (2.29). Substituting the value of $\frac{\partial}{\partial\beta} l^*(\beta; \mathbf{X},\ \mathbf{Y})$ from (2.34) in the left hand side of equation (2.29) we get:

$$\begin{aligned} E^*\left(\frac{\mathbf{W}^t R^{-1}}{\sigma_v^2}\{\mathbf{Y}-\mathbf{W}\beta\} + \frac{tr(R^{-1})\Lambda\beta}{\sigma_v^2}\right) &= E^*\Bigg(\frac{\mathbf{W}^t R^{-1}\mathbf{Y}}{\sigma_v^2} - \frac{\mathbf{W}^t R^{-1}\mathbf{W}\beta}{\sigma_v^2} \\ &\quad + \frac{tr(R^{-1})\Lambda\beta}{\sigma_v^2}\Bigg) \\ &= \Bigg(\frac{\mathbf{X}^t R^{-1}\mathbf{Y}}{\sigma_v^2} - \frac{\{\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X} + tr(R^{-1})\Lambda\}\beta}{\sigma_v^2} \\ &\quad + \frac{tr(R^{-1})\Lambda\beta}{\sigma_v^2}\Bigg) \\ &= \frac{1}{\sigma_v^2}\Bigg[\mathbf{X}^t R^{-1}\mathbf{Y} - \mathbf{X}^t R^{-1}\mathbf{X}\beta\Bigg] \\ &= \frac{\mathbf{X}^t R^{-1}}{\sigma_v^2}(\mathbf{Y}-\mathbf{X}\beta) \end{aligned} \tag{2.37}$$

Moreover, from Equation (2.27), we have:

$$\begin{aligned} \frac{\partial}{\partial\beta} l^*(\beta; \mathbf{X},\ \mathbf{Y}) &= \frac{\partial}{\partial\beta}\left[g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2}\left\{\mathbf{Y}^t\mathbf{R}^{-1}\mathbf{Y} - 2\beta^t\mathbf{X}^t\mathbf{R}^{-1}\mathbf{Y} + \beta^t\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X}\beta\right\}\right] \\ &= \frac{1}{\sigma_v^2}\left[\mathbf{X}^t\mathbf{R}^{-1}\mathbf{Y} - \mathbf{X}^t\mathbf{R}^{-1}\mathbf{X}\beta\right] \\ &= \frac{1}{\sigma_v^2}\mathbf{X}^t\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{X}\beta), \end{aligned} \tag{2.38}$$

which is the same as Equation (2.37). Hence the required condition in Equation (2.29) holds. Since the measurement error is in the covariates $\mathbf{X}$ and E^* is the conditional expectation with respect to $\mathbf{W}$ given $\mathbf{V}$ and $\mathbf{Y}$, the correction for the likelihood $l(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ will be same. The required conditions can also be checked along the same lines as $l(\beta; \mathbf{W},\ \mathbf{Y})$. Hence the corrected likelihood function $l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ shall be given by:

$$\begin{aligned} l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) = g(\sigma_v^2, \Sigma) - \frac{1}{2\sigma_v^2}\Bigg[&\mathbf{Y} - \mathbf{W}\beta - \mathbf{IV}]^T[\mathbf{Y} - \mathbf{W}\beta - \mathbf{IV}] \\ &- tr(R^{-1})\beta^T\Lambda\beta\Bigg] - \frac{1}{2\sigma_v^2}\mathbf{V}^T\Sigma^{-1}\mathbf{V}, \end{aligned} \tag{2.39}$$

and the corrected score function, using result in Equation (2.21) for partial derivatives is obtained as:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{V}} l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) &= \frac{1}{\sigma_v^2}[\mathbf{Y} - \mathbf{W}\beta - \mathbf{I}\mathbf{V}] - \frac{\Sigma^{-1}\mathbf{V}}{\sigma_v^2} \\ &= \Psi^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}).\end{aligned} \tag{2.40}$$

2.5.1. Estimation of β and V

For obtaining the consistent estimates we have to check that the corrected score functions are unbiased. Let us consider the corrected score function for β first. If E is the global expectation, we have:

$$E\{\Psi^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})\} = E^+[E^*\{\Psi^*(\beta, \mathbf{W},\ \mathbf{Y})\}]. \tag{2.41}$$

Substituting from Equation (2.35), we have:

$$\begin{aligned} E\{\Psi^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})\} &= E^+\left[E^*\left\{\frac{\mathbf{W}^t R^{-1}}{\sigma_v^2}(\mathbf{Y} - \mathbf{W}\beta) + \frac{1}{\sigma_v^2} tr(R^{-1})\Lambda\beta\right\}\right] \\ &= E^+\left[E^*\frac{\mathbf{W}^t R^{-1}\mathbf{Y}}{\sigma_v^2} - E^*\frac{\mathbf{W}^t R^{-1}\mathbf{W}\beta}{\sigma_v^2} + \frac{tr(R^{-1})\Lambda\beta}{\sigma_v^2}\right] \\ &= E^+\Bigg[\frac{\mathbf{X}^t R^{-1}\mathbf{Y}}{\sigma_v^2} - \frac{\{\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X} + tr(R^{-1})\Lambda\}\beta}{\sigma_v^2} \\ &\qquad + \frac{tr(R^{-1})\Lambda\beta}{\sigma_v^2}\Bigg] \\ &= E^+\left\{\frac{\mathbf{X}^t R^{-1}\mathbf{Y}}{\sigma_v^2} - \frac{(\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X})\beta}{\sigma_v^2}\right\} \\ &= \frac{\mathbf{X}^t R^{-1} E^+(\mathbf{Y})}{\sigma_v^2} - \frac{(\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X})\beta}{\sigma_v^2} \\ &= 0, \end{aligned} \tag{2.42}$$

since $E^+(\mathbf{Y}) = \mathbf{X}\beta$. Thus the corrected score function is unbiased and it will yield consistent estimator of β. Along similar lines, we can also show that the corrected score function $\Psi^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ satisfies the condition (2.29) and is unbiased. Hence the equations:

$$\Psi^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) = 0 \tag{2.43}$$

$$\Psi^*(\beta; \mathbf{W},\ \mathbf{Y}) = 0, \tag{2.44}$$

shall produce consistent estimators. From Equation (2.44) we have:

$$\frac{\mathbf{W}^t R^{-1}}{\sigma_v^2}(\mathbf{Y} - \mathbf{W}\beta) + \frac{1}{\sigma_v^2} tr(R^{-1})\Lambda\beta = 0$$

$$\mathbf{W}^t R^{-1}\mathbf{Y} - \mathbf{W}^t R^{-1}\mathbf{W}\beta + tr(R^{-1})\Lambda\beta = 0$$

$$\{\mathbf{W}^t R^{-1}\mathbf{W} - tr(R^{-1})\Lambda\}\beta = \mathbf{W}^t R^{-1}\mathbf{Y}.$$

Hence the corrected scores estimator of β is given by:

$$\hat{\beta} = \{\mathbf{W}^t R^{-1}\mathbf{W} - tr(R^{-1})\Lambda\}^{-1}\mathbf{W}^t R^{-1}\mathbf{Y}. \tag{2.45}$$

Similarly, from Equation (2.43) we have:

$$\Psi^*(\hat{\beta}; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) = 0$$

$$\frac{1}{\sigma_v^2}\{\mathbf{Y} - \mathbf{W}\hat{\beta} - \mathbf{IV}\} - \frac{\Sigma^{-1}\mathbf{V}}{\sigma_v^2} = 0.$$

This yields the corrected scores estimator of $\mathbf{V}$ as:

$$\hat{V} = (I + \Sigma^{-1})^{-1}(\mathbf{Y} - \mathbf{W}\hat{\beta}), \tag{2.46}$$

where $\hat{\beta}$ is given by equation (2.45).

CHAPTER III

EFFICIENT SMALL AREA ESTIMATION WHEN COVARIATES ARE MEASURED WITH ERROR USING SIMULATION EXTRAPOLATION

3.1. Introduction

In this chapter we use a simulation based bias correction technique called simulation extrapolation to obtain estimates of regression coefficients in the Fay-Herriot model when covariates are measured with error. This bias correction method has simplicity, generality and approximate-inference characteristics. Simulation extrapolation is ideally suited to problems with additive measurement error, and more generally to any problem in which the measurement error generating process can be imitated on a computer via Monte Carlo methods.

As Buzas *et al.* (2005) remark, SIMEX is a simulation-based method estimating and reducing bias due to measurement error. SIMEX estimates are obtained by adding additional measurement error to the data in a resampling-like stage, establishing a trend of measurement error-induced bias versus the variance of the added measurement error, and extrapolating this trend back to the case of no measurement error. The technique was proposed by Cook and Stefanski (1994), further developed by Cook and Stefanski (1995), Carroll, Kuchenhoff, Lombard, and Stefanski (1996), Devanarayan (1996), Carroll and Stefanski (1997) and Devanarayan and Stefanski (2002). SIMEX is closed related to the Monte Carlo corrected score (MCCS) method considered in Chapter IV.

The fact that measurement error in a predictor variable induces bias in parameter estimates is counterintuitive to many people. An integral component of SIMEX is a self-contained simulation study resulting in graphical displays that illustrate the effect illustrate the effect of measurement error on parameter estimates and the need for bias correction. SIMEX is very general in the sense that the bias due to measurement error in almost any estimator of almost any parameter is readily estimated and corrected, at least approximately. In the absence of measurement error, it is often the case that competing estimators are available that are consistent for the same parameter and only differ asymptotically with respect to sampling variability. However, these same estimators can be differentially affected by measurement error.

The key idea underlying SIMEX is the fact that the effect of measurement error can be determined experimentally via simulations. In a study of the effect of radiation exposure on tumor development in rats, for example, one is naturally led to an experiment in which radiation dose is varied. Similarly, in a study of the biasing effects of measurement error on an estimator, one is naturally led to an experiment in which the level of measurement error is varied. So if we regard measurement error as a factor whose influence on an estimator is to be determined, we consider simulation experiments in which the level of the measurement error, as measured by its variance, is intentionally varied.

Usually the measurement error variance is known or can be reasonably well estimated, say from replicate measurements, as will be exhibited through a real data example in this chapter. Cook and Stefanski (1994) remark that SIMEX combines the features of a parametric bootstrap and method of moments inference. They said that it is simple to implement and applicable to a broad range of settings. Because of its ability to lend itself to graphic description, it is particularly useful in situations where the user, though familiar with the standard statistical methods, wants to steer clear of burdensome technical details of model fitting and statistical theory that generally accompanies all but the simplest of measurement error models. Because the method is completely general, it is also useful in applications when the general method under consideration is novel and may not have been thoroughly studied and developed. Assessing the variability in the estimates is, however a challenging task, so we have opted to use the bootstrap or jack-knife estimates to assess sampling variability, as was originally suggested by Cook and Stefanski (1994). They developed this method in response to the need for fitting nonstandard generalized linear measurement error models, specifically models in which the mean function depends on something other than a linear function of the predictor measured with error. Before SIMEX came into being, the general approaches described by Stefanski (1985), Fuller (1987, Chapter 3), Whittemore and Keller (1988), and Carroll and Stefanski (1990) could be applied to such models, but besides their complexity and the need for specialized software, these methods entailed approximations, the quality of which would be difficult to verify with complicated models.

Why is there a need for bias correction in measurement error models? Cook and Stefanski (1994) remark that although the attenuating effects of measurement error have been well-documented and publicized in literature, it often happens that before a measurement error analysis is undertaken, it is necessary to explain the need for such an analysis to non-statisticians. When corrections for attenuation are made, they are often regarded with skepticism above and beyond the normal amount that should accompany any statistical modeling. The reason is that corrections for attenuation are often in the researcher's best interest. For example, assuming that the *presence* of an effect has been convincingly demonstrated, the perceived importance of the effect is likely to depend on its magnitude. A correction for attenuation generally increases the magnitude of the effect; thus it is in the researcher's interest to make such corrections. For example in health economics, consider a situation wherein a pharmaceutical company is estimating the cost-effectiveness of an anti-hypertensive drug. A correction for attenuation due to measurement error in blood pressure readings produces a greater predicted risk for a given reduction in blood pressure (MacMahon *et al.* 1990). Consequently the drug's cost-effectiveness is enhanced. Clearly, a correction for attenuation is in the company's best interest and is likely to be viewed by skepticism by anyone not versed in the field of measurement error modeling.

In such situations just described, there is a need for corrections for attenuation that are both statistically and scientifically defensible while also being demonstratively conservative. The SIMEX method is capable of producing estimates with the desired properties.

The outline of this chapter is as follows. In Section 3.1, we describe the SIMEX algorithm as suggested by Cook and Stefanski and then discuss how it could be applied to model with homoscedastic errors with known error variance, heteroscedastic errors with known error variances and heteroscedastic errors with unknown variances and replicate measurements.

In Section 3.2, we discuss how this SIMEX method is applied to the Fay Herriot model for small area estimation where covariates are measured with error. We study the asymptotic properties of the estimators of the regression coefficients in the model. We also study the rate of convergence of the composite estimator of population total in the i^{th} small area when the SIMEX estimates are used. Jackknife estimates of the mean squared error are also suggested.

In Section 3.3, we describe a simulation study in which it is shown, by means of the empirical mean squared error and Jackknife estimates of mean squared error, that the SIMEX estimator is superior to the Lohr-Ybarra estimator with respect to efficiency when either all or a fraction of the small areas have covariates measured with error. We also study the performance of the SIMEX estimator and Lohr-Ybarra estimators when departure from normality of measurement error occurs, specifically for errors that have a Student's t or a skew normal distribution. In the Section 3.4.1 we examine the performance of these estimators for a multiplicative error model along the lines of the methods suggested by Carroll *et al.* (2006) and mention a method suggested by Lechner (2007) for tackling multiplicative measurement error.

In Section 3.4, we use the SIMEX estimator in a real data set obtained from the 2003-2004 National Center for Health Statistics, U.S. National Health and Nutrition Examination Survey and use data from the 2004 National Center for Health Statistics, U.S. National Health Interview Survey (NHIS) as auxiliary information. Jackknife estimates of the mean squared errors were calculated and the superiority of the SIMEX estimators over direct estimators and the Lohr-Ybarra estimators is thus exhibited. As Lohr and Ybarra (2008) had pointed out, there is a significant increase in efficiency and estimated precision for small domains when use of auxiliary information is made through the measurement error model versus ignoring availability of auxiliary information altogether.

3.2. The Simulation-Extrapolation Correction

It is easy to see that SIMEX is fully consistent for linear models and approximate for nonlinear models. The algorithm demonstrates that SIMEX is founded on the observation that bias in parameter estimation varies in a systematic way with the magnitude of the measurement error. Details of the method are best understood in the context of the classical additive measurement error model. However, the method is not limited to this model. The SIMEX algorithm can be summarized as follows:

- In the simulation step, *pseudo-errors*, that are independent measurement errors and have a variance of ζC_i, are generated and added to the original auxiliary data X_i thereby creating data sets with successively larger measurement error. As in Lohr and Ybarra (2008) we consider C_i's to be known, but for applications, replace them with precise estimates to be described later in the chapter. The quantity ζ describes the different levels of pseudoerror added to the data, and its values are usually taken from 0 to 2 by the SIMEX R package.

- A *re-measurement* of the auxiliary data $\widehat{X}_i$ is done and a new *pseudo-variable* $\widetilde{X}_i$ is defined. The MSE of the pseudo-variable is an increase over the MSE of the original auxiliary information $\widehat{X}_i$ by a multiplicative factor of $(1+ \zeta)$.

- Next, estimates are obtained from each of the generated, contaminated data sets in each area i.

- The simulation and estimation steps are repeated for a large number of times and the average value of the estimate for each level of contamination (different values of ζ) is calculated. These averages are plotted against the ζ values and a regression technique, for example non-linear least squares is used to fit an extrapolant function to the averaged, error-contaminated estimates.

- Extrapolation to the ideal case of no pseudo-measurement error ($\zeta = -1$) yields the SIMEX estimate.

It has been seen, and will be demonstrated via simulations that minor violations of the assumption of normality of the measurement errors is not critical in practice. We assume that the measurement error variance, C_i for each small area is known or can be estimated sufficiently well to be regarded as known.

SIMEX is applicable to general estimation procedures, for example, least-squares, weighted least squares (the method used for estimating the regression coefficients by Lohr and Ybarra), maximum likelihood, quasilikelihood, etc. While proving the asymptotic properties of the SIMEX estimator when applied to the Fay-Herriot model, we will not distinguish among the methods, but instead refer to the *estimator* to be the chosen estimation method computed as if there were no measurement error, i.e., the naive estimate.

In our model, the re-measured pseudo- variables for the b^{th} iteration ($b = 1, ..., B$) are,

$$\widetilde{X}_{b,i} = \widehat{X}_i + \sqrt{\zeta} U_{b,i}. \tag{3.1}$$

So the $MSE(\widetilde{X}_{b,i}) = (1+\zeta)C_i$. Hence we see the MSE has increased by a factor of $(1+\zeta)$. The value of the SIMEX estimate of β is the one obtained by extrapolating to $\zeta = -1$.

A key property of simulated pseudo-data is as follows: Since $E(\widetilde{X}_{b,i}|X_i) = X_i$ (since by assumption, $E(\widehat{X_i}) = X_i$ and the U_i's have an expected value of zero) it means that the MSE of $\widetilde{X}_{b,i}$ as a re-measurement of X_i, defined as $E\{(\widetilde{X}_{b,i}(\zeta) - X_i)^2/X_i\}$

converges to 0 as $\zeta \to -1$. The reason for averaging over many simulations is that we are interested in estimating the extra bias due to added measurement error, not in inducing more variability, and averaging reduces the Monte Carlo simulation variation. Note that although we cannot add measurement error with negative variance, $\zeta C_i = -C_i$, we can add measurement error with positive variance, determine the form of the bias as a function ζ, and extrapolate to the hypothetical case of adding negative variance ($\zeta = -1$).

The SIMEX estimator thus obtained is used in the composite estimator for the population total Y_i in i^{th} small area, similar to one suggested by Lohr and Ybarra. We now proceed to describe how the SIMEX algorithm can be modified for heteroscedastic errors and replicate measurements.

3.2.1. Heteroscedastic Errors with Known Error Variances

This is the model that Ybarra and Lohr suggested, namely, they supposed that we auxiliary information $\hat{X}_i = X_i + U_i$, where U_i has variance C_i, is independent of X_i, Y_i and v_i and C_i is known for all i. This is not a common error model but it provides an important error model, but it provides a useful stepping stone to other, more common heteroscedastic error models. It is an appropriate model to use when $\hat{X}_i$ is the mean of $k_i \geq 1$ replicate measurements, each having a known variance C, in which case $C_i = C_i/k_i$.

In this case the only change in the algorithm is that the remeasurement procedure in Equation (3.1) is replaced by

$$\widetilde{X}_{b,i} = \widehat{X}_i + \sqrt{\zeta} U_{b,i}, \tag{3.2}$$

for $i = 1, ..., m$ and $b = 1, ..., B$. Here, the pseudo errors, $U_{b,i}$ are again mutually independent, independent of all observed data, and identically distributed random variables with mean zero and variance C_i. Note that

$$var\{\widetilde{X}_{b,i}|X_i\} = (1+\zeta)C_i = (1+\zeta)Var(\hat{X}_i|X_i) \tag{3.3}$$

and $E(\widetilde{X}_{b,i}/X_i) = X_i$. So we see that the two variances $var\{\widetilde{X}_{b,i}|X_i\}$ and $var\{\hat{X}_i|X_i\}$ differ by a multiplicative factor that vanishes when $\zeta = -1$ and consequently $MSE\{\widetilde{X}_{b,i}\}$ $= E[(\widetilde{X}_{b,i} - X_i)^2|X_i] \to 0$ as $\zeta \to -1$, the key property of the remeasured data. The ensuing steps are described before.

3.2.2. Heteroscedastic Errors with Unknown Variances and Replicate Measurements

SIMEX can also be used in a model that allows for arbitrary unknown heteroscedastic error variances. SIMEX for this model was developed and studied by Devanarayan (1996) and Devanarayan and Stefanski (2002). For this model $k_i \geq 2$ replicate measurements are necessary for each subject in order to identify error variances C_i. The assumed error model is $\widetilde{X}_{ij} = X_i + U_{ij}$ where $j = 1, ..., k_i$ are assumed Normal(0, C_i) independent of X_i, v_i, and Y_i with all the C_i's *unknown*. With replicate measurements, the best measurement of X_i is the mean $\bar{X}_{i.}$ and we define the so-called naive estimation procedure as doing the usual, nonmeasurement error analysis, of the data $(Y_i, \bar{X}_{i.})$.

Because the variances C_i's are unknown, we cannot generate remeasured data as before in Equation (3.1). However, recall that the key property of the remeasured data is that the variance of the best measurement of X_i is inflated by the factor $1+\zeta$. With replicate measurements, we can obtain such variance-inflated measurements by taking suboptimal linear combinations of the replicate measurements. This is done by using random linear contrasts.

Suppose that $\mathbf{l}_{b,i} = (l_{b,i,1}, ..., l_{b,i,k_i})^T$ is a normalized contrast vector, $\sum_j l_{b,i,j} = 0$ and $\sum_j l_{b,i,j}^2 = 1$. Define

$$\widetilde{X}_{b,i} = \widehat{\bar{X}_i} + \left(\frac{\zeta}{k_i}\right)^{1/2} \sum_{j=1}^{k_i} l_{b,i,j} \hat{X}_{i,j} \tag{3.4}$$

for $i = 1, ..., m$ and $b = 1, ..., B$. With this definition, a little calculation shows that $E(\widetilde{X}_{b,i}/X_i) = X_i$ and

$$var\{\widetilde{X}_{b,i}|X_i\} = (1+\zeta)C_i/k_i = (1+\zeta)var(\widehat{\bar{X}_i}|X_i). \tag{3.5}$$

Thus the remeasurements in this model too have the same key property as the ones before, that is, the variances of the error in the remeasurements are inflated by a multiplicative factor that vanishes when $\zeta = -1$ and MSE of $\widetilde{X}_{b,i}$ converges to 0 as $\zeta \to -1$.

Because we want to average over B remeasured data sets, we need a way to generate random, replicate versions of (3.4). We do this by making contrasts random. We get statistical replicates of $\widetilde{X}_{b,i}(\zeta)$ by sampling $\mathbf{l}_{b,i}$ uniformly from the set of all normalized contrast vectors of dimension k_i. This is accomplished using pseudorandom Normal(0,1) random variables. If $Z_{b,i,1}$,...,Z_{b,i,k_i} are Normal(0,1), then

$$l_{b,i,j} = \frac{Z_{b,i,j} - \bar{Z}_{b,i,.}}{\sqrt{\sum_{j=1}^{k_i}(Z_{b,i,j} - \bar{Z}_{b,i,.})^2}}, \tag{3.6}$$

are such that $\sum_j l_{b,i,j} = 0$ and $\sum_j l_{b,i,j}^2 = 1$. Furthermore, the random contrast vector $\mathbf{l}_{b,i} = (l_{b,i,1}, ..., l_{b,i,k_i})^T$ is uniformly distributed on the set of all normalized contrast vectors of dimension k_i (Devanarayan and Stefanski, 2002).

The averaged naive estimates and then the SIMEX estimator are then obtained in the same way as before. Because this version of SIMEX generates pseudo-errors from the observed data (via random contrasts), we call it empirical SIMEX to distinguish if from versions of SIMEX that generate pseudo errors from a parametric normal model.

3.3. Properties

In this section we look at some of the theoretical properties of the SIMEX-based estimator we suggest. All technical details of the proof are given in the Appendix A to this chapter.

<u>*Result 1:*</u> Assume that the Lohr-Ybarra model described in Chapter II in Equation (2.6) holds, that $MSE(\widehat{X}_i) = C_i$ and that $(\widehat{X}_i, v_i, e_i)$ is independent of $(\widehat{X}_j, v_j, e_j)$ when $i \neq j$. Within area i, assume that v_i, X_i and e_i are mutually independent. Let $\widetilde{Y}_{iME} = \gamma_i y_i + (1-\gamma_i)\widehat{X}_i^T\beta$, where:

$$\gamma_i = \frac{\sigma_v^2 + \beta^T C_i \beta}{\sigma_v^2 + \beta^T C_i \beta + \psi_i}$$

Then Lohr and Ybarra (2008) showed that $\widetilde{Y}_{iME}$ has the minimum mean squared error amongst all linear combinations of y_i and $\widehat{X}_i^T\beta$ of the form $a_i y_i + (1-a_i)\widehat{X}_i^T\beta$. In addition,

$$E(\widetilde{Y}_{iME} - Y_i) = (1-\gamma_i)\{E(\widehat{X}_i) - X_i^T\}\beta. \tag{3.7}$$

They showed that the mean squared error of $\tilde{Y}_{iME}$ is

$$MSE(\widetilde{Y}_{iME}) = \gamma_i \psi_i. \tag{3.8}$$

Remark 1: Lohr and Ybarra (2008) noted that $\gamma_i = MSE(r_i)/MSE(r_i) + \psi_i$ depends on the error incurred in estimating X_i as well as on ${\sigma_v}^2$ and ψ_i. $\widetilde{Y}_{iME}$ relies most heavily on the regression estimator (or in our case, the SIMEX estimator of β) when $\hat{X}_i$ is measured without error. In that case, $\widetilde{Y}_{iME}$ reduces to the univariate Fay-Herriot estimator. If $\hat{X}_i$ is measured imprecisely, then $\widetilde{Y}_{iME}$ depends more heavily on the direct estimator y_i. If y is not measured in area i then $\gamma_i = 0$, $\widetilde{Y}_{iME} = \hat{X^T}_i\beta$ and $MSE(\widetilde{Y}_{iME}) = \sigma_v^2 + \beta^T C_i \beta$.

Remark 2: Lohr and Ybarra (2008) also pointed out that the measurement error predictor has mean squared error at most as large as that of the direct estimator since γ_i lies between 0 and 1. The predictor $\tilde{Y}_{iME}$ is thus an improvement over the predictor that ignores measurement error altogether, which could perform worse than the direct estimator y_i. If auxiliary information from a lot of the small areas, that is, the X_i's are measured accurately, there is great improvement in efficiency.

Result 2: If any estimator $\hat{\beta}$ of any parameter β is asymptotically normal when X_i is observed, then the corresponding SIMEX estimator $\widehat{\beta}_{SIMEX}$ is also asymptotically normal. Furthermore assume that the conditions on Result 1 hold. Assume that $\hat{X}_i$ has mean X_i and covariance matrix C_i. Suppose that $\hat{X}_i$, v_i and e_i are independent with uniformly bounded $3+\eta$ moments for some $\eta > 0$ for $i = 1, ..., m$. Also, suppose that the elements of X_i, C_i as well as the positive constant ψ_i are uniformly bounded for $i = 1, ..., m$ then,

$$\hat{\beta}_{SIMEX} \approx \beta + O_p(m^{-1/2}) \tag{3.9}$$

Remark 1: Our measurement error model assumes that independence between measurement errors and other variables in the data sets exists. Cook and Stefanski (1994) commented that although the IID pseudoerrors are generated under these assumptions, the simulated sets of errors have non-zero correlations with Y_i, $\hat{X}_i$, X_i, nonzero sample means and sample variances $\neq 1$. They said that the effects of these random departures from expected behavior can be eliminated by generating the pseudoerrors to be uncorrelated with the observed data and normalized to have sample mean and variance equal to zero and one respectively. They called the NON-IID pseudoerrors. They also showed that the asymptotic bias in $\hat{\beta}_{SIMEX}$ is of the same order of magnitude for the quadratic and non-linear extrapolant function. Carroll *et al.* (1996) showed that for NON-IID errors, the SIMEX estimator is of the order $O_p(mB)^{-1/2}$

Remark 2: In our model we take m and B as fixed, but Carroll *et al.* (1996) investigated three types of asymptotics for the SIMEX estimator: (a) $m \to \infty$ for B fixed, (b) $m \to \infty$ and $B \to \infty$ simultaneously, and (c) $B = \infty$ and $m \to \infty$ showed that when B. In all these cases they showed that the same results for asymptotic normality hold.

Result 3: Under certain regularity conditions, in the Lohr-Ybarra model, $\hat{\beta}_{SIMEX}$ is an approximately consistent estimator of β.

Result 4: If all the conditions and properties specified in results (1), (2) and (3) hold then,

$$\hat{\sigma}_v^2 = \frac{1}{(m-p)} \sum_{i=1}^{m} \{(y_i - \hat{X}_i^T \hat{\beta}_{SIMEX})^2 - \psi_i - \hat{\beta}_{SIMEX}^T C_i \hat{\beta}_{SIMEX}\} \tag{3.10}$$

is a consistent estimator of $\hat{\sigma}_v^2$.

Remark 1: The estimator in result 4 is similar in form to the estimator of σ_v^2 suggested by Lohr and Ybarra (2008), but they use a weighted least squares estimator of β which we replace with a SIMEX estimator. Lohr and Ybarra, in turn, borrowed this methodology for estimation σ_v^2 from the one suggested by Prasad and Rao (1990). However, Prasad and Rao adjust for the $O(m^{-1})$ leverage in the regression and have $C_i = 0$ in their application.

Remark 2: Lohr and Ybarra (2008) commented that one consequence of using this estimator of σ_v^2 is that it has greater variance than the corresponding estimator when the X_i's are known. In their simulations, the extra variability caused a higher fraction of the estimates $\hat{\sigma}_v^2$ to be set equal to zero, resulting in inflated variances for the jack-knife estimates of $MSE(\tilde{Y}_{iME})$. Using a SIMEX estimator instead of a weighted least squares estimator for β compensates for this loss in precision caused by negative variance estimates, as will be shown through simulation. Lohr and Ybarra suggest using a restricted maximum likelihood estimator (REML) like suggested in Datta and Lahiri (2000), using only observations with small measurement error. A Bayesian estimator of σ_v^2, using an inverse gamma prior might exhibit more stability and is being currently investigated. A more detailed discussion of the options avaliable for estimation of variance components is given in Section 3.5 of this chapter.

Result 5: Assume that the conditions and properties specified in result (1), (2), (3), and (4) hold, that the elements of C_i and X_i are uniformly bounded. Let $\hat{\Theta} = (\hat{\sigma}_v^2, \hat{\beta}_{SIMEX})^T$ with,

$$cov(\hat{\Theta}) = \begin{bmatrix} var(\hat{\sigma}_v^2) & s_m^T \\ s_m & D_m \end{bmatrix} + o(m^{-1}) = O(m^{-1}), \tag{3.11}$$

where $s_m = cov(\widehat{\sigma}_v^2, \hat{\beta}_{SIMEX})^T$ and $D_m = var(\widehat{\beta}_{SIMEX})$. Assume that $\widehat{\Theta}$ is independent of $(\hat{X}_i, y_i)$ and that the sixth central moments of Θ are $o(m^{-1})$. Then

$$\begin{aligned} MSE(\widehat{Y}_{iSIMEX}) &= \gamma_i\psi_i + (1-\gamma_i)^2 tr\{(C_i + X_i X_i^T)D_m\} \\ &\quad + \frac{\psi_i^2}{(\beta^T C_i \beta + \sigma_v^2 + \psi_i)^3} E\Big\{\widehat{\sigma}_v^2 + \widehat{\beta}_{SIMEX}^T C_i \widehat{\beta}_{SIMEX} \\ &\quad - \sigma_v^2 - \beta^T C_i \beta\Big\}^2 + 2E\{(1-\widehat{\gamma}_i)^2(\widehat{\beta}_{SIMEX} - \beta)^T\} C_i \beta \\ &\quad + o(m^{-1}). \end{aligned} \tag{3.12}$$

Remark 1: Lohr and Ybarra (2008) noted that if $C_i = 0$, result (5) gives the mean squared error derived in Prasad and Rao (1990) for the univariate Fay-Herriot estimator. The extra terms in the mean squared error due to the presence of measurement error in $\widehat{X}_i$ are of the same order, $O(1/m)$ as the g_{2i} and g_{3i} given in Prasad and Rao (1990).

Remark 2: Lohr and Ybarra (2008) noted that the mean squared error may be estimated by analytically obtaining estimators of the terms in Equation (3.12) that are unbiased to the order $o(1/m)$, as done in Prasad and Rao (1990), Datta and Lahiri (2000) and Datta *et al.* (2005) for the Fay-Herriot estimator. In this method, estimators are substituted for D_m and the expected values in (3.6). Additional terms in the estimated mean squared error will vary depending upon the specific estimator used for Θ (Datta *et al.*, 2005).

3.4. Simulation Study

A simulation study was conducted after using SIMEX for bias correction in the above model. We generated of $X_i \sim N(5, 9)$, $\psi_i \sim$ Gamma(4.5, 2). For each iteration we generated $Y_i = 1 + 3x_i + v_i$, $y_i = Y_i + e_i$ and $\widehat{x}_i = x_i + u_i$, where v_i, e_i and u_i are independent normal variables with mean 0 and variance σ_v^2, ψ_i and c_i respectively.

Consider 3 factors (Lohr and Ybarra, 2008) : Factor 1: $\sigma_v^2 = 2$, 3 or 4; Factor 2: $c_i \in \{0, d\}$ for d= {2,3, or 4}; Factor 3: m= 20, 50 or 100. The number of iterations for each combination were 10000. Three different scenarios w.r.t. X_i were considered, i.e., ALL of them being measured with error (k=100), some (specified percentage k) measured with error and NONE ($k = 0$) of them measured with error. SIMEX estimates were thus obtained after generating pseudo-variables.

We found empirical MSE's, for each area i, for the direct, Fay-Herriot estimator that ignores measurement error and treats $\widehat{X}_i$ as the actual observed X_i, Lohr-Ybarra and SIMEX estimators. These empirical MSE's are calculated as, $\Sigma_{l=1}^{10000}\Sigma_j(\widehat{Y}_{i(l)} - Y_{i(l)})^2/10000$ where $Y_{i(l)}$ and $\widehat{Y}_{i(l)}$ are the true and predicted values of $X_i^T\beta + v_i$ in l^{th} iteration. For the estimated weights $\widehat{\gamma}_{iv}$, the consistent estimator $\widehat{\sigma}_v^2$ provided by Lohr-Ybarra is used and while calculating $\widehat{Y}_{iSIM}$ the SIMEX estimator of β is substituted in (3.4) in place of the weighted least squares estimator of β. Such an estimator of σ_v^2 will be *approximately consistent* due to the approximate consistency of the SIMEX estimator of β, which is proved later. The results for $m = 100$, $c_i = 3$, $\sigma_v^2 = 4$ are displayed in Table 1:

Table 1. Empirical mean squared error for the four estimators, y_i, $\widetilde{Y}_{iS}$, $\widehat{Y}_{iME}$, $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, measurement error variance $C_i = 3$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	c_i	y_i	$\widetilde{Y}_{iS}$	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	0	8.9	3.3	3.2	3.4
20	3	10.4	6.9	7.8	3.2
50	3	9.2	4.3	7.4	3.2
80	3	9.8	6.7	7.0	5.6
100	3	11.8	5.4	5.6	5.0

Table 2. Absolute value of the bias for the four estimators, y_i, $\widetilde{Y}_{iS}$, $\widehat{Y}_{iME}$, $\widehat{Y}_{iSIMEX}$ when the number of small areas is 50, measurement error variance $C_i = 2$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	y_i	$\widetilde{Y}_{iS}$	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	1.41	0.20	0.23	0.21
20	1.62	0.43	0.68	0.25
50	1.56	0.76	0.47	0.37
80	1.72	0.84	0.52	0.24
100	1.93	0.86	0.67	0.31

In the Tables 1 and 2, four estimators are compared. y_i is the direct estimator, $\widetilde{Y}_{iS}$ that blindly uses the Fay-Herriot estimator $\widetilde{Y}_{iS} = \widehat{\gamma}_{iv} y_i + (1 - \widehat{\gamma}_{iv})\widehat{X}_i^T \beta$ in which $\widehat{\gamma}_{iv}$ and $\widehat{\beta}$ are calculated assuming $\widehat{X}_i$'s are the true values, $\widehat{Y}_{iME}$ is the Lohr-Ybarra estimator and $\widehat{Y}_{iSIM}$ is our SIMEX estimator. In Table 2, absolute value of the bias is reported for the estimators when m=50, C_i= 2 ; $\sigma_v^2 = 4$ and k as before. It can be seen that the direct estimator y_i performs the worst almost for all cases, i.e., these estimates have very large average empirical mean squared errors and hence should be used with care. The SIMEX estimate is the best choice when auxiliary information in a majority of the small areas is measured with error. As the proportion of areas having auxiliary information measured with error increases, the empirical mean squared error of the SIMEX estimator is seen to be lower than that of the other estimators.

When we use the model that uses the true covariates, it is but obvious that $\widehat{Y}_{iFH}$, the Fay-Herriot estimator consistently performs the best. But the fallacy of ignoring measurement error in covariates and using the Fay-Herriot estimator even when the exact X_i have not been observed, is brought to light by the increase in the mean squared error and bias of $\widetilde{Y}_{iS}$ as k increases.

We propose that the Jack-knife estimates for deriving the mean squared errors of population total estimates derived by Jiang *et al.* (2002) should also be used and the performance of the SIMEX estimates be compared.

This jackknife estimate is of the form $\widehat{MSE}(\widehat{Y}_{iSIM}) = \widehat{M}_{1i} + \widehat{M}_{2i}$, where:

$$\widehat{M}_{1i} = \widehat{\gamma}_i \psi_i + \frac{m-1}{m} \sum_{j=1}^{m} (\widehat{\gamma}_i \psi_i - \widehat{\gamma}_{i(-j)} \psi_i), \tag{3.13}$$

where the notation $(-j)$ indicates an estimator of the same form but based on the data set without area j. Also:

$$\widehat{M}_{2i} = \frac{m-1}{m} \sum_{j=1}^{m} (\widehat{Y}_{iSIMEX(-j)} - \widehat{Y}_{iSIMEX})^2, \tag{3.14}$$

where,

$$\widehat{Y}_{iSIMEX(-j)} = \widehat{\gamma}_{i(-j)} y_i + (1 - \widehat{\gamma}_{i(-j)}) \widehat{X_i} \widehat{\beta}_{SIMEX(-j)}. \tag{3.15}$$

In Table 3, jackknife estimates of mean squared error for the Lohr-Ybarra and SIMEX estimators are given for m=50, C_i= 2 ; $\sigma_v^2 = 4$. The jackknife estimates of mean squared error also highlight the better performance of the SIMEX estimators as the proportion k increases, in comparison to the Lohr-Ybarra estimator. This superior performance of the SIMEX estimator persists, regardless of the fact that the number of small areas is small, as is evident from Table 4.

Table 3. Jackknife estimates of the mean squared error of the Lohr-Ybarra estimator $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, measurement error variance $C_i = 2$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	3.5	3.4
20	4.6	3.3
50	5.7	3.4
80	5.4	3.2
100	7.5	3.1

Table 4. Jackknife estimates of the mean squared error of the Lohr-Ybarra estimator $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is relatively small, i.e., 20, measurement error variance $C_i = 3$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	5.5	5.4
20	5.6	5.3
50	6.6	4.7
80	6.7	4.4
100	7.2	4.2

Table 5. Absolute value of bias for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, $\sigma_v^2 = 4$ and the U_i are generated from a t-distribution with 10 degrees of freedom. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	1.23	1.22
20	1.31	1.06
50	1.36	0.64
80	1.57	0.62
100	1.93	0.75

3.4.1. Studying Departure from Normality of Measurement Error

A simulation study was also conducted by generating the measurement error U_i from a heavy-tailed distribution like t-distribution with 10 and 5 degrees of freedom and skew normal distribution with skewness parameter equal to -1 and 2 . The absolute value of the bias was calculated for the Lohr-YBarra estimator and SIMEX estimators. The results of this study of the performance of the SIMEX estimates of Y_i's are displayed in Tables 5, 6, 7 and 8.

Table 6. Empirical mean squared error for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 50, $\sigma_v^2 = 4$ and the U_i are generated from a t-distribution with 5 degrees of freedom. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	5.87	5.86
20	5.34	5.21
50	6.68	5.20
80	6.73	4.79
100	6.74	4.72

Table 7. Absolute value of bias for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, $\sigma_v^2 = 2$ and the U_i are generated from a skew-normal distribution with location parameter 0, scale parameter 2 and skewness parameter -1 (left-skewed distribution). k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	1.34	1.33
20	2.32	1.25
50	1.76	1.04
80	1.77	0.98
100	1.83	0.77

Table 8. Empirical mean squared error for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, $\sigma_v^2 = 2$ and the U_i are generated from a skew-normal distribution with location parameter 0, scale parameter 3 and skewness parameter 2 (right-skewed distribution). k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$
0	6.04	6.05
20	5.85	5.35
50	5.78	5.34
80	5.79	4.98
100	6.28	4.87

From the results in Tables 5, 6, 7 and 8, it can be concluded that SIMEX estimators are robust to departure from normality of the measurement error. The bias and mean squared errors for the measurement errors generated from a t-distribution with 5 degrees of freedom may seem unsatisfactorily high, but even so, the SIMEX estimator is still superior to the Lohr-Ybarra estimator. As k (the proportion of small areas having auxiliary information measured with error) increases, the bias of the SIMEX estimator is consistently lower than that of the Lohr-Ybarra estimator. We investigated the behavior of the SIMEX estimator when the U_i's are generated from a skew-normal distribution and even though generation of errors from this distribution increases the bias and mean squared errors greatly in comparision to errors from normal or t distributions, the SIMEX estimator still outperforms the Lohr-Ybarra estimator, irrespective of whether the measurement error distribution is right or left skewed. R.B. Arellano-Valle *et al.* (2005) considered a skew-normal measurement error model and then proceeded to a Bayesian analysis based on a special class of prior distributions yielding invariance in terms of the fact that the posterior distribution is in the class of the ordinary normal distribution and also developed an EM-type algorithm which can overcome some difficulties detected by using direct maximization of the likelihood. Carroll *et al.* (1999) used Bayesian MCMC and Gibbs Sampler to deal with the problem of a mixture-normal measurement error model.

3.4.2. Studying Departure from Additivity of Measurement Error Model

Another problem that is widely encountered while analyzing data with measurement error is the non-additivity of the measurement error model. Multiplicative measurement error has the advantage that the original observation is proportionally modified so that a single observation with a higher value, which is typically subject to a higher disclosure risk, is better masked as through additive measurement error. Furthermore, a multiplicative measurement error conserves the structural zeros contained in the data set.

Less attention has been given in the literature to multiplicative measurement error. Hwang (1986) derives a consistent estimator for the slope parameter in a linear regression model in the presence of multiplicative measurement error, by correcting the asymptotic bias of the least squares estimator. He however treated the X_i's as random variables with an unknown distribution, and not as fixed like in the Fay-Herriot model. The only assumption that he makes about the measurement error is that it is independent and identically distributed. Lin (1989) extends the model of Hwang (1986) to include measurement error in both the dependent and the independent variables, and shows that the derived modified OLS estimator is consistent under the assumptions that the multiplicative measurement error takes only positive values, and is independent and identically distributed. Iturria, Carroll, and Firth (1999) consider a polynomial regression model in the presence of multiplicative measurement error. They compare two methods differing in their assumptions about the distribution of the measurement error, and the distribution of the unobserved variable. They derive a consistent estimator and asymptotic standard error using moment estimation methods.

The assumption of the additivity measurement error may be unreasonable in certain data sets. However, SIMEX applies more generally and is often extended to other error models. Eckert, Carroll and Wang (1997) considered measurement error of the form $h(\hat{X}_i) = h(X_i) + U_i$ where $h(\cdot)$ is a monotone transformation function selected from some family H of monotone functions. The idea of their model is that, in the correct scale, measurement error is additive. One of the models they suggested was the multiplicative error model $\hat{X}_i = Xe^{U_i}$ which give additivity in the logarithmic scale, that is, $log(\hat{X}_i) = log(X_i) + U_i$. This is the same principle we use and compare how the SIMEX estimator of Y_i behaves in the additive and multiplicative measurement error case.

Carroll *et al.* (2006) mentioned that by using the logarithm transformation and converting it to an additive model, the remeasured data (after adding pseudoerrors) are:

$$log\{\hat{X}_{b,i}(\zeta)\} = log(\hat{X}_i) + \sqrt{\zeta}U_{b,i},$$

where $U_{b,i}$ are $N(0, C_i)$ pseudorandom variables. Note that upon transformation:

$$\hat{X}_{b,i}(\zeta) = exp\{log(\hat{X}_i) + \sqrt{\zeta}U_{b,i}\}, \tag{3.16}$$

In the additive error model, the key property of the remeasured data was the fact that variance was increased by the multiplicative factor $1+\zeta$ and that this multiplier vanishes when $\zeta = -1$. The multiplicative model has a similar property, but the relevant measure here is mean squared error and not variance (since $E(\hat{X}_i|X_i) \neq X_i$). Carroll *et al.* (2006) stated that,

$$MSE\{\hat{X}_{b,i}(\zeta)|X_i\} = k^*(\zeta, C_i)MSE(\hat{X}_i|X_i)$$

where,

$$k^*(\zeta, C_i) = \frac{\{e^{C_i(1+\zeta)} - 1\}^2 + e^{C_i(1+\zeta)}\{e^{C_i(1+\zeta)} - 1\}}{(e^{C_i} - 1)^2 + e^{C_i}(e^{C_i} - 1)},$$

is such that $k^*(0, C_i) = 1$, $k^*(\zeta, C_i)$ is increasing in $\zeta > 0$ for all C_i, $\lim_{\zeta \to -1} k^*(\zeta, C_i)$ is 0. Thus, this can be treated as the biased error model counterpart of the one with additive measurement error. Ordinary least squares estimators of β are computed at each value of ζ, and these estimates are used as before in the convex estimator of Y_i.

Remark: Lechner (2007) commented that statistical offices collect many types of micro data which contain highly sensitive information, whose confidentiality has to be protected against disclosure in the interest of the observational unit itself, but also in the interest of the data collecting institutions. Variables which contain sensitive information are for example those which include geographic details like the region where a firm is settled, the sector where a firm is active, and extreme values, for example the profit or the turnover of a large firm. Obviously, for the data collecting institutions, this creates a trade-off between the goal of guaranteeing confidentiality and the provision of providing the maximum amount of information to the researcher. Therefore, they become interested in the creation of scientific use files which optimally combine the interests of the survey respondents, of the data collecting institutions and of the academic users. However, this dual vision of the problem is quite recent, and for many years, the statistical institutions have restricted their interest only on protecting the data, without focusing on the potential usability of the anonymized data. Multiplicative measurement error, is more suitable to protect data against disclosure. Multiplicative measurement error has the advantage that the original observation is proportionally modified so that a single observation with a higher value, which is typically subject to a higher disclosure risk, is better masked as through additive measurement error.

Table 9. Absolute value of bias for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ calculated ignoring the multiplicative error, the SIMEX estimator $\widehat{Y}^{add}_{iSIMEX1}$ when additive error exists, the SIMEX estimator $\widehat{Y}^{add}_{iSIMEX2}$ when there is multiplicative measurement error but we ignore it and proceed like there is additive error present, the SIMEX estimator $\widehat{Y}^{mult}_{iSIMEX}$ when we consider the multiplicative error and apply the logarithmic transformation, when the number of small areas is 60, $\sigma_v^2 = 3$ and the U_i are generated from a $Normal(0, 4)$ distribution. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}^{add}_{iSIMEX1}$	$\widehat{Y}^{add}_{iSIMEX2}$	$\widehat{Y}^{Mult}_{iSIMEX}$
0	1.53	1.54	1.55	1.54
20	2.42	1.53	1.54	1.61
50	1.86	1.44	1.52	1.65
80	1.87	1.38	1.55	1.71
100	1.93	1.21	1.52	1.68

Table 10. Empirical mean squared errors for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ calculated ignoring the multiplicative error, the SIMEX estimator $\widehat{Y}^{add}_{iSIMEX1}$ when additive error exists, the SIMEX estimator $\widehat{Y}^{add}_{iSIMEX2}$ when there is multiplicative measurement error but we ignore it and proceed like there is additive error present, the SIMEX estimator $\widehat{Y}^{mult}_{iSIMEX}$ when we consider the multiplicative error and apply the logarithmic transformation, when the number of small areas is 50, $\sigma_v^2 = 2$ and the U_i are generated from a $Normal(0, 3)$ distribution. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}^{add}_{iSIMEX1}$	$\widehat{Y}^{add}_{iSIMEX2}$	$\widehat{Y}^{Mult}_{iSIMEX}$
0	3.43	3.44	3.44	3.45
20	3.52	2.57	2.54	2.91
50	2.89	2.74	2.71	2.87
80	3.27	2.88	2.84	3.75
100	3.28	2.21	2.42	3.48

The results in Tables 9 and 10 clearly show that the Lohr-Ybarra estimator $\widehat{Y}_{iME}$ calculated ignoring the multiplicative error and the SIMEX estimator $\widehat{Y}^{add}_{iSIM2}$ calculated when there is multiplicative measurement error but we ignore it and proceed like there is additive error present- both of them portray a dangerous picture since they give a misleading notion of precision. But it is unwise to proceed in such a manner when multiplicative error exists since the estimates obtained in such a fashion will be highly inconsistent. The SIMEX estimator $\widehat{Y}^{add}_{iSIM1}$ obtained when additive error exists is, though, still slightly superiror to $\widehat{Y}^{mult}_{iSIM}$, the SIMEX estimator when we consider the multiplicative error and apply the logarithmic transformation. All in all, even though the SIMEX estimator seems to perform better in the presence of additive measurement error, it is still a better option than the Lohr-Ybarra estimator when measurement error is multiplicative. These conclusions are corroborated in the Table 10 of empirical mean squared errors.

Remark: As Hwang (1986) points out, the first thought of a statistician when he sees a multiplicative measurement error model is usually to apply a logarithmic transformation, so that the multiplicative measurement error model becomes an additive one. However the problem of this approach is that the logarithmic transformation destroys the linearity of the model, so that the common methods developed. Lechner (2007) suggested that logarithmic transformation destroys the linearity of the model, so that the common methods developed to correct the effects of measurement error on the properties of linear estimators generally fail. She suggested that the SIMEX method has to be modified in order to be applied to the case of multiplicative measurement error, without using a logarithmic transformation. The approach she suggested involves the assumption that $E(X_i|U_i) = 1$ and she assessed the performance of her modification of the SIMEX method using the linear, quadratic and non-linear extrapolant function. The application of her suggested modification in the Fay-Herriot model with measurement error in covariates is the subject of current research.

3.5. Estimation of Variance Components

Rao (2003b) mentions that the basic area level model has a limitation, namely, the assumption of known sampling variances ψ_i's is restrictive, although methods based on generalized variance functions (GVF) have been proposed to produce smoothed estimates of ψ_i's. Fay and Herriot (1979) suggest a variance stabilizing transformation of the data to a scale in which the sampling variances could be considered known, identification and similar transformation of auxiliary variables available for each small area, derivation of a composite estimator for the parameter of interest for each small area and finally the re-transformation of the resulting estimates to the original scale. However, this procedure of re-transformation may result in the loss of precision in estimates. Hence, various methods of estimating sampling variances ψ's have been proposed.

Approach 1: Rivest and Vandal (1999) studied the effect of estimating ψ_i on mean squared error of the empirical Bayes estimator with $\widehat{\gamma}_i = \widehat{\sigma}_v^2/(\widehat{\sigma}_v^2 + \widehat{\psi}_i)$ where $\widehat{\psi}_i$ is an estimator of ψ_i. For example, suppose that we have a random sample $y_{ij} \sim N(\theta_i, \widehat{\sigma}^2)$, $j = 1, ..., n_i$. from i^{th} small area and $\widehat{\theta}_i = \bar{y}_i$, the sample mean. In this case, $\psi_i = S_i^2/n_i$ is design-unbiased for ψ_i, where S_i^2 is the sample variance. Further, $\bar{y}_i$ and $\widehat{\psi}_i$ are independently and normally distributed.

Approach 2: Lohr and Ybarra (2008) have suggested using REML estimates for σ_v^2 since they are much more stable. Datta and Lahiri (2000) used ML and REML estimators of σ_v^2 and ψ_i and obtained a second order approximation of the MSE of the estimator $\widehat{\theta}_i(\widehat{\psi}_i)$ of Y_i. In their approximation they assume that m is large and neglect all terms of order $o(m^{-1})$.

Approach 3: Of all the estimators of ψ_i suggested in recent times, the most pertinent and relevant to our model (one with measurement error in covariates) is the one suggested by Torabi *et al.* (2009). They consider a unit level model, i.e.,

$$y_{ij} = \beta_0 + \beta_1 X_i + v_i + e_{ij} \tag{3.17}$$

where $j = 1, ..., N_i$, $i = 1, ..., m$ and in reality, $X_{ij} = x_{ij} + U_{ij}$ are observed. N_i is the known population size of the i^{th} small area. y_{ij} is the value of the study variable associated with the j^{th} unit in the i^{th} small area and X_i is the unknown true area-specific covariate associated with the y_{ij}. Random errors e_{ij}, measurement errors U_{ij} and area-level random effects v_i are mutually independent with $U_{ij} \sim N(0, \sigma_u^2)$, $v_{ij} \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. They considered the structural measurement error model where $X_i \sim N(\mu_x, \sigma_x^2)$ are independent of e's, v's and U's. Hence they denote the vector of model parameters by $\mathcal{D} = (\beta_0, \beta_1, \mu_x, \sigma_v^2, \sigma_e^2, \sigma_u^2, \sigma_x^2)^T$. A sample of size n_i is selected from i^{th} small area.

The consistent estimator for the elements of $\mathcal{D}$ they suggested were:

$$\widehat{\sigma}_u^2 = MSW_x = \frac{SSW_x}{n_T - m} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{n_T - m},$$

where $n_T = \sum_{j=1}^m n_i$ is the total sample size and $\bar{X}_i = 1/n_i \sum_{j=1}^{n_i} X_{ij}$.

$$\widehat{\sigma}_e^2 = MSW_y = \frac{SSW_y}{n_T - m} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_T - m},$$

where $\bar{y}_i$ is the sample mean of y_{ij}'s in the i^{th} small area.

$$\widehat{\beta}_1 = [(MSB_x - MSW_x)(m-1)]^{-1} \sum_{i=1}^{m} n_i \bar{y}_i (\bar{X}_i - \bar{X}),$$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{X},$$

$$\widehat{\mu}_x = \bar{X},$$

where $\bar{X} = {n_T}^{-1} \sum_{i=1}^{m} n_i \bar{X}_i$, $\bar{y} = {n_T}^{-1} \sum_{i=1}^{m} n_i \bar{Y}_i$ and
$MSB_x = (m-1)^{-1} \sum_{i=1}^{m} n_i (\bar{X}_i - \bar{X})^2$

$$\widehat{\sigma}_x^2 = max\left\{0, \frac{(MSB_x - MSW_x)(m-1)}{g_m}\right\},$$

and

$$\widehat{\sigma}_x^2 = max\left\{0, \frac{m-1}{g_m}(MSB_y - MSW_y) - \widehat{\beta}_1^2 \widehat{\sigma}_x^2\right\}$$

where $MSB_y = (m-1)^{-1} \sum_{i=1}^{m} n_i (\bar{y}_i - \bar{y})^2$ and $g_m = n_T - \frac{\sum {n_i}^2}{n_T}$.

The area-level model analog of these estimates in the unit-level model when there is measurement error in covariates, is the subject of current research. We propose to use the SIMEX estimator of β and then compare the efficiency and consistency of these ANOVA-type estimates of the variance components to that of the REML estimates suggested in approach 2.

3.6. Data Example

Like Lohr and Ybarra (2008), we apply the SIMEX method to a data set from the 2003-2004 National Center for Health Statistics, U.S. National Health and Nutrition

Examination Survey (NHANES), using the 2004 National Center for Health Statistics, U.S. National Health Interview Survey as auxiliary information. The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. The small areas in this case are 30 demographic subgroups cross-classified by race and ethnicity, namely Mexican American, Non-Hispanic Black and Non-Hispanic White; by age group, namely 20-39, 40-59, 60 years and above and by gender. Confidentiality issues greatly influenced the criteria for selecting the small areas. The race and ethnicity categories used here are the most detailed currently released for public use on the CDC (Center for Disease Control and Prevention) website.

Height and weight for each respondent are measured National Health and Nutrition Examination Survey medical examination by government interviewers and the body mass index (BMI) is calculated as $height/weight^2$. In contrast to this method, in the National Health Interview Survey, the body mass index is calculated using responses reported by the subjects themselves through a questionnaire. There is a strong case for using a measurement error model for the NHIS responses since it has been seen that respondents might not report their heights and weights accurately. Hence the National Health Interview Survey may not be measuring the actual body mass index but a variable that is highly related to it. Hence the body mass index calculated in the National Interview Health Survey is a natural choice for the auxiliary variable in this study. As is seen in the figure on page 72 and as expected the body mass index values from the National Health Interview Survey and the National Health and Nutrition Examination Survey are highly related, with a correlation coefficient of 0.95. The 2003-2004 National Center for Health Statistics, National Health and Nutrition

Examination Survey has body mass index values for 4424 persons in the demographic subgroups of interest, with domain sample sizes ranging from 10 to 358. The National Health Interview Survey has a sample size for reported body mass index was around 30,000 with domain sample sizes between 90 to 3090. For confidentiality, only area level data was released for public use, i.e. only the average body mass index for each subgroup was available along with the standard error of the averages.

To estimate measurement error variance, data collected on body mass index in the previous years was used, treating data from consecutive National Health Interview Survey responses as replicates. The estimates $\widehat{Y}_{iME}$ (Lohr-Ybarra estimator), $\widehat{Y}_{iSIM}$ (SIMEX estimator) of the mean body mass index in each of the 30 small areas, and the jackknife estimates of their mean squared errors were calculated in R version 2.13. For $\widehat{X}_i^T = (1, x_i)$, the SIMEX method produces the estimate of the regression coefficient as 0.95 and intercept as 0.16. The corresponding estimate of σ_v^2 using the SIMEX method was 0.52. Thus the Lohr-Ybarra seems to be slightly underestimating the slope and intercept terms, but overestimating the variance of the random effects.

The figure on page 73 plots the jacknife estimates of the mean squares errors of both estimators and it is clearly demonstrated that while the SIMEX estimators contains the mean squared errors, using the Lohr-Ybarra method can produce inflated values, as high as 5.6. Clearly, SIMEX performs as a highly efficient method of bias correction in this case. It is worth noting here that the performance of the direct estimator y_i was also assessed and it performs the *worst* in comparison to the Lohr-Ybarra estimator and the SIMEX estimators. Thus if auxiliary information is available it should be made use of to improve the efficiency of small area estimates. But if we ignore the error in the auxiliary information and proceed to calculate the Fay-Herriot estimates pretending that the $\hat{X}_i$'s are the true population quantities, then the estimated mean squared errors might give the deceptive appearance of precision, but they are not consistent estimators for the true mean squared errors.

The better performance of the SIMEX estimators in the study can surely be attributed to the existence of slight departure from normality in the data set, specially for the auxiliary information obtained from the NHIS study. This departure from normality is evident in the figure on page 74 of the boxplots for the NHIS and NHANES body mass index values. We have already shown via simulation that the SIMEX estimator performs much better than Lohr-Ybarra estimator in the presence of extreme departures from normality. So the former's superiority in practical applications such as these is to be expected.

Remark: Lohr and Ybarra treat the measurement error variances and the regression model error (e_i) variances as known and do not mention in their paper how they estimated them while applying their method to this data set. So we used the same values of the estimates of these variances while calculating the Lohr-Ybarra estimates of Y_i as we did for the SIMEX estimators. They might have used a different method of variance component estimation or had access to additional information about them. In any case, this might be a possible reason for any difference between our Lohr-Ybarra estimates and the ones given in their paper. It is important to note that this points to another shortcoming in their method- that of its impracticality, since it is not always plausible to treat these variances as known, and even if one chooses to do so, precise estimates must be suggested for application purposes.

Figure 1 clearly shows that the NHIS body mass index values are a wise choice for an auxiliary variable since there is high correlation between them and the NHANES body mass index values.

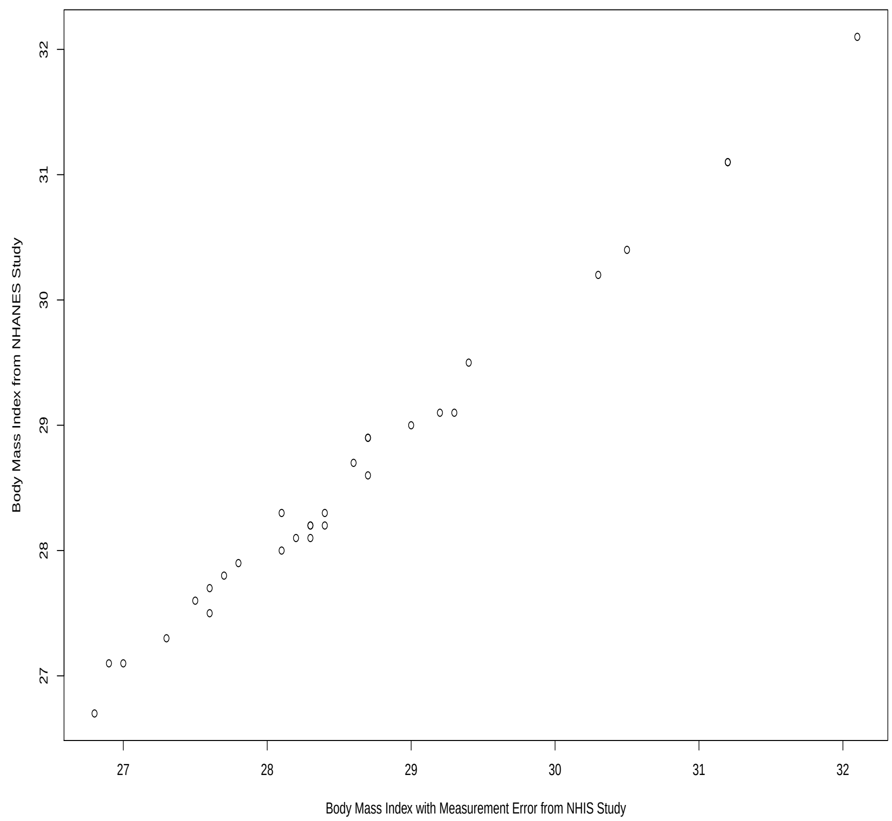

Fig. 1. Plot of mean body mass index from the National Health and Nutrition Examination Survey (NHANES) against reported body mass index from the National Health Interview Survey (NHIS) which has measurement error for 30 domains.

The superior performance of the SIMEX estimator compared to the Lohr-Ybarra estimator is thus quite evident from Figure 2. The boxplots in Figure 3 point towards the non-normal nature of the extent and might be a possible explanation for such a superior performance by the SIMEX estimator in this case.

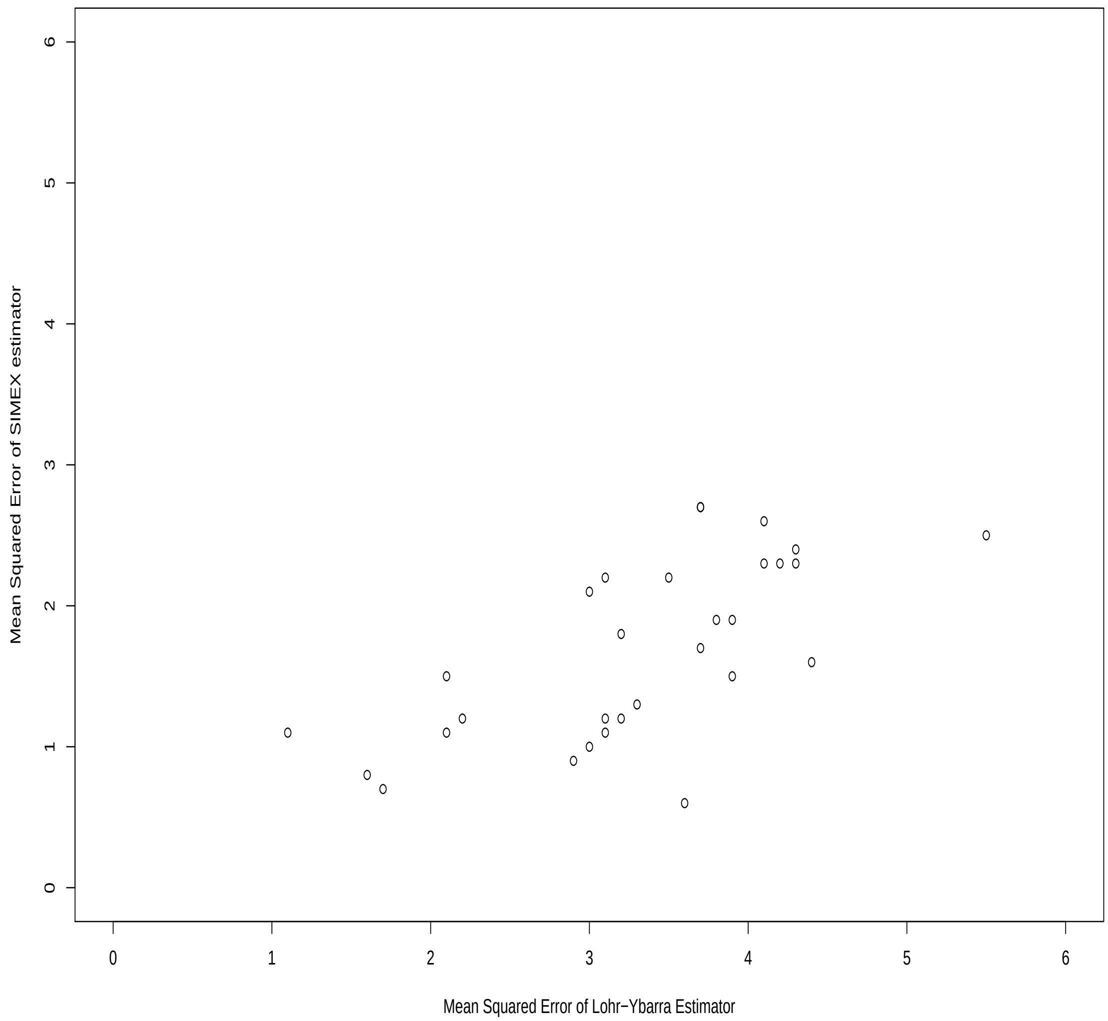

Fig. 2. Plot of jackknife estimates of mean squared error of the Lohr-Ybarra estimator ($\hat{Y}_{iME}$) of the mean body mass index, against the jackknife estimates of mean squared error of the SIMEX estimator ($\hat{Y}_{iSIM}$) for the 30 domains or small areas.

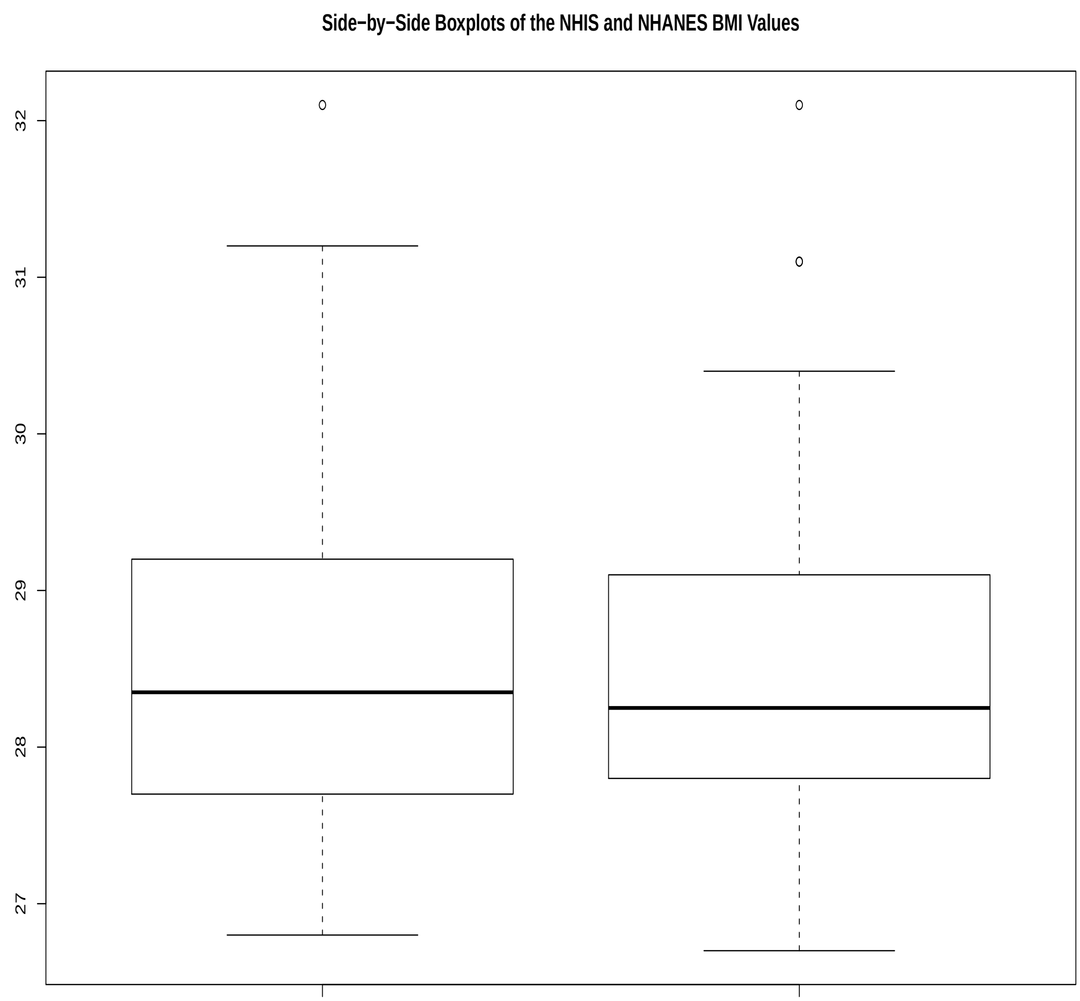

Fig. 3. Side-by-side boxplots of the body mass index values obtained from the National Health Interview Survey (NHIS) which we treat as auxiliary information and the body mass index values from the National Health and Nutrition Examination Survey (NHANES) which is the response variable in the study.

CHAPTER IV

EFFICIENT SMALL AREA ESTIMATION WHEN COVARIATES ARE MEASURED WITH ERROR USING CORRECTED SCORES

Statistical models, in our case the Fay-Herriot model, whose independent variables are subject to measurement errors are often referred to as "errors-in-variables models". To correct for the effects of measurement error on parameter estimation, this chapter considers a correction for score functions. A corrected score function is one whose expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true independent variables. This approach makes it possible to do inference as well as estimation of model parameters without additional assumptions.

Simulation extrapolation is a widely applicable method as has been demonstrated in Chapter I. It produces estimators that are consistent in important special cases, such as linear regression and loglinear mean models, but for more complicated, general models, they are only approximately consistent. Corrected scores is a method that results in fully consistent estimators more generally. Although our model is not a complicated one, with respect to assumptions of linearity of the model and normality of errors and random effects, corrected scores lends itself very easily to extensions to more complex models. Consistency is achieved by virtue of the fact that the estimators are M-estimators whose score functions are unbiased in the presence of measurement error. This property is also true of structural model maximum likelihood and quasilikelihood estimates as discussed in Carroll *et al.* (Chapter 8, 2006). The lack of assumptions pertaining to the unknown X_i distinguishes the methods based on corrected scores from the others.

We do not deal with structural modeling for our model, that is we treat the X_i's as fixed parameters, but a word of caution about corrected scores and functional modeling is necessary. With additive, normally distributed measurement error, functional maximum likelihood maximizes the joint density of the observed data with respect to all of the unknown parameters, including the X_i's. While this works perfectly with linear regression (and hence, in our Fay-Herriot model), it fails for more complex models such as logistic regression (Carroll and Stefanski, 1987b). The functional estimator in most nonlinear models is both extremely difficult to compute and not even consistent or valid. Buzas *et al.* (2005) remarked that regression parameter estimators in nonlinear models are defined implicitly through estimating equations. Estimating equations are often based on the likelihood score, i.e. the derivative of the log-likelihood, or quasi-likelihood scores that only require assumptions on the first and second conditional moments of the model. The method of corrected scores is also applicable to the case of least squares equations and also leads to parameter estimation based on estimating equations.

When independent variables or covariates in models are subject to error, ordinary estimators for model parameters are often inconsistent as is demonstrated by Hwang (1986), Anderson and Gerbing (1984), Fuller (1980) for linear models; Armstrong (1985) for generalized linear models; Carroll and Stefanski (1987a) and Wolter and Fuller (1982) for nonlinear models; Carroll and Stefanski (1987b) for logistic regression models; Prentice (1982) for Cox regression models. Fuller (1987) provides a good introduction and fundamental aspects mainly for linear models. Stefanski (1985) presents a general method of constructing an estimate with smaller bias. Armstrong (1985) proposes a modified GLIM procedure accounting for covariate measurement error which is particularly suited to the generalized linear model; an obstacle to the broader applicability of Armstrong's proposed method is that it requires information such as the posterior distribution of true covariates given observed covariate values. It is sometimes possible even in nonlinear models to deal with an expanded likelihood equation which incorporates the measurement error distribution (Wolter and Fuller, 1982). However, this approach often needs additional assumptions such as, assuming that the true covariates are random samples from a certain distribution or that the magnitude of the measurement error approaches zero as the sample size increases.

Corrected scores deals with a method in which the score function itself is corrected so that it yields consistent estimates. The idea is that the conditional distribution of the corrected estimate given the true independent variables and the dependent variables is centered around the maximum likelihood estimate, which in turn is centered around the true parameter value. As one of the consequences, the consistent estimates corrected for measurement errors cannot have smaller variance than the maximum likelihood estimate without measurement errors. An advantage of this approach over Stefanski's theory (1985) is that it finds the regression coefficient corrected for attenuation in the linear model and it determines more precise variances of the estimates. This method is applied to the Fay-Herriot model, which is a special case of a linear mixed model. Carroll and Stefanski (1987a) deal with an unbiased score function in generalized linear models from a completely different approach. In short, the corrected score function we suggest is an unbiased score function, but the converse is not necessarily true. It is not always possible to obtain a corrected score (Stefanski, 1989). For example, the likelihood score for logistic regression does not admit a corrected score, except under certain restrictions (Buzas and Stefanski, 1996). Methods for obtaining corrected scores and approximately corrected scores via computer simulation have been developed (Novick and Stefanski, 2002; Devanarayan and Stefanski, 2002).

Novick and Stefanski (2002) noted that the corrected-score method, first proposed by Nakamura (1990) is attractive because of its generality and its relationship to method-of-moments estimation. It is a functional method and thus possesses theoretical robustness properties that structural model methods, which make distributional assumptions about the true predictors, do not have. One obstacle to using the corrected-score method is the determination of the corrected score. These have been identified for particular models (Nakamura, 1990; Carroll *et al.*, 1995) on a model-by-model basis, although the results in Gray *et al.* (1973) provide one approach to obtaining corrected scores as infinite series. For many complex problems, a corrected score can be obtained via Monte Carlo simulation, referred to as Monte Carlo corrected scores (MCCS).

Gimenez and Bolfarine (1997) remarked that Nakamura's approach to estimation in measurement error models is based on a corrected score function, and the claim that the estimators obtained are consistent for functional models. Proof of the claim essentially assumed the existence of a corrected log-likelihood for which differentiation with respect to model parameters can be interchanged with conditional expectation taken with respect to the measurement error distributions, given the response variables and true covariates. There is a rich literature dealing with simple yet practical models for which this assumption is false, i.e. a corrected score function for the model may not be obtained through differentiating a corrected log-likelihood although it exists like in the case of a logistic regression model. Novick and Stefanski Monte Carlo approach is applicable more generally.

The outline of this chapter is as follows. In Section 4.1, we describes score-function based methods like corrected scores and conditional scores used in measurement error modeling. We describe Novick and Stefanski's Monte Carlo corrected score approach and point out its link to simulation extrapolation. Then in Section 4.2, we proceed to show that exact corrected scores do exist in the case of the Fay Herriot model with normal measurement error and derive corrected estimators for the regression coefficients β's and also random effects v_i's.

In Section 4.3, we study the properties of the corrected scores estimators in our model and show that they are asymptotically normal. The rate of convergence of the estimators in also studied in detail.

In Section 4.4, we describe an algorithm for estimation of variance components and show how it is applied in the case of the Fay-Herriot model with measurement error in covariates. In Section 4.5, we describe a simulation study conducted to compare the efficiency of our corrected scores estimators to the Lohr-Ybarra estimator, conditional scores estimator, Monte Carlo corrected scores estimator and SIMEX estimator. Like in Chapter II, the robustness of the corrected scores estimators is also studied via simulation.

In Section 4.6, we apply the corrected scores approach to the data set used and described in Chapter III to obtain estimates of the mean body mass index for certain demographic subgroups. The jacknife estimates of the mean squared errors of these estimates are calculated and compared to those of the Lohr-Ybarra and SIMEX estimates mentioned in Chapter III.

4.1. What are Corrected Scores?

In the functional modeling of measurement error, the distribution of the error-prone covariate X_i (also sometimes referred to as the latent variable) is not specified, the methods based on estimating equations are widely used in practice. Functional modeling is appealing in the situations where we do not have much knowledge about the behavior of the covariates. The estimating equation methods used in functional modeling include the conditional score function method proposed by Carroll and Stefanski (1987a) and corrected score function method proposed by Stefanski (1989) and Nakamura (1990).

The two broad classes of score function methods that are frequently used, as described by Carroll *et al.* (2006), are:

- The corrected score method effectively estimates the estimator that one would use if there was no measurement error in the covariates.
- The conditional-score method that exploits special structures in important models such as linear, logistic, Poisson, loglinear, and gamma-inverse regression, using a traditional statistical device, conditioning on sufficient statistics, to obtain estimators.

A general description of the corrected scores approach can be given as follows. Let Y be the continuous response variable, and X and Z be the vectors of covariates and β be the vector of parameters to be estimated. Let Z be the covariate measured without error and let X be measured with non-differential additive classical error. That is, $W = X + U$ is observed instead of X, U being the additive error whose distribution is specified. Thus the observed data are (Y_i, W_i, Z_i) and not (Y_i, X_i, Z_i)

for $i = 1, 2, ..., m$.

An estimating equation $\Psi_i(\beta; Y_i, W_i, Z_i)$ is called an unbiased estimating equation for β if it satisfies,

$$E\{\Psi_i(\beta; Y_i, W_i, Z_i)\} = 0, \tag{4.1}$$

for $i = 1, 2, ..., m$. These equations may be obtained either by using likelihood or the least squares theory of estimation. Here, m is the sample size, which in our case becomes the number of small areas in the study being conducted. The function Ψ_i is the score function from the model for the data without error.

Generally, if X_i is replaced by its surrogate variable W_i, the estimating equation $\Psi_i(\cdot)$ is no longer unbiased. The measurement error introduces bias into $\Psi_i(\cdot)$ and this bias, in turn, translates into the bias in the estimates of the coordinates of β.

An unbiased estimating function leads to a consistent estimator for β under the regularity conditions required for the consistency of maximum likelihood estimates. That is, as $m \to \infty$, the solution $\widehat{\beta}_{CS}$ to:

$$\sum_{i=1}^{m}\{\Psi_i(\beta; Y_i, W_i, Z_i)\} = 0, \tag{4.2}$$

converges in probability to the true value of β. When W_i is observed instead of X_i, the naive estimation function $\Psi_i(Y_i, W_i, Z_i)$ is no longer unbiased. However, if a modified (or corrected) version $\Psi_i^*(Y_i, W_i, Z_i)$ is unbiased under expectations with respect to W_i, conditional on the true data (Y_i, X_i, Z_i). then solving,

$$\sum_{i=1}^{m}\{\Psi_i^*(\beta; Y_i, W_i, Z_i)\} = 0, \tag{4.3}$$

will yield a consistent estimator for β.

It suffices to construct $\Psi_i^*(\beta; Y_i, W_i, Z_i)$ such that,

$$E^*_{W/Y,X,Z}\{\Psi_i^*(\beta; Y_i, W_i, Z_i)\} = \Psi_i(\beta; Y_i, X_i, Z_i). \tag{4.4}$$

This means that $\Psi_i^*(\beta; Y_i, W_i, Z_i)$ is an unbiased estimator of $\Psi_i(\beta; Y_i, X_i, Z_i)$ under conditional expectations with respect to W_i given the true data (Y_i, X_i, Z_i) and hence it is referred to as the corrected score function.

The corrected score functions are unbiased whenever the original scores are. That is, if E is the global expectation such that $E = E^+E^*$ where E^* is the conditional expectation with respect to W_i given (Y_i, X_i, Z_i) and E^+ is the expectation with respect to the Y_i considering X_i and Z_i as fixed. Then,

$$\begin{aligned} E\{\Psi_i^*(\beta; Y_i, W_i, Z_i)\} &= E^+_Y[E^*_{W/Y,X,Z}\{\Psi_i^*(\beta; Y_i, W_i, Z_i)\}] \\ &= E^+_Y\{\Psi_i(\beta; Y_i, X_i, Z_i)\} \\ &= 0, \end{aligned}$$

since from above, $\Psi_i^*(\beta; Y_i, W_i, Z_i)$ is an unbiased estimator for $\Psi_i(\beta; Y_i, W_i, Z_i)$ and the last expectation is zero due to the assumption that the original score $\Psi_i(\beta; Y_i, W_i, Z_i)$ is unbiased. The unbiasedness of the score function is needed to guarantee the consistency of the estimators obtained from the corrected score functions. However, an unbiased score function is not necessarily a corrected one. The consistency of $\widehat{\beta}_{CS}$ will be verified assuming that the parameter β is identifiable. The condition for identifiability is given in Proposition 1 of Nakamura's paper (1990) and will be verified for our Fay-Herriot model later in this chapter.

Nakamura (1990) has proved that the solution of corrected score functions, if they exist, provide consistent estimators under certain regularity conditions. He showed that the exact corrected score functions exist for some regression models in the general linear model family such as Gaussian, Poisson, Gamma, inverse-Gaussian and Wald regression models. For logistic regression model, however, an exact corrected score does not exist.

Remark: Nakamura has shown that if $\Psi_i^*(\beta; Y_i, W_i)$ is a corrected score function, then β^* obtained by solving $\sum_{i=1}^{m}\{\Psi_i^*(\beta; Y_i, W_i)\} = 0$ is a consistent estimator of β. Its asymptotic distribution is normal with mean β and variance-covariance matrix is given by:

$$\Sigma_{CS} = \frac{1}{m}\{A_{CS}^{-1}\}\{B_{CS}\}\{A_{CS}^{-T}\}, \tag{4.5}$$

where,

$$\begin{aligned} A_{CS} &= -E\left\{\frac{\partial}{\partial\beta}\Psi^*(\cdot)\right\} \\ B_{CS} &= E\{\Psi^*(\cdot)\Psi^*(\cdot)^T\} \\ A_{CS}^{-T} &= (A_{CS}^{-1})^T \end{aligned}$$

4.1.1. Monte Carlo Corrected Scores (MCCS)

As mentioned before, exact corrected scores are not always available for certain complex models, such as the logistic regression model, for which Novick and Stefanski (2002) developed a simulation-based approach to obtain corrected scores. This approach is known to yield fully consistent estimators more widely than regression calibration or SIMEX. This is because these estimators are M-estimators whose score functions are unbiased in the presence of measurement error. Corrected scores approach works in a way similar to SIMEX, i.e., we generate via simulation pseudo-variables that define an estimator.

This method works as follows: consider a linear regression model with mean $E(\mathbf{Y}/\mathbf{Z},\mathbf{X}) = \beta_o + {\beta_z}^t\mathbf{Z} + {\beta_x}^t\mathbf{X}$, $var(\mathbf{Y}/\mathbf{Z},\mathbf{X}) = \sigma^2$ and the classical additive, non-differential error model $\mathbf{W} = \mathbf{X} + \mathbf{U}$ with $U \sim N(0, \Sigma_{uu})$, where Σ_{uu} is known. Let the unknown regression parameters be $\Theta_1 = (\beta_o, {\beta_z}^t, {\beta_x}^t)^t$ and $\Theta = ({\Theta_1}^t, {\Theta_2}^t)^t$ with $\Theta_2 = \sigma^2$

The ordinary least squares score function for multiple linear regression in the absence of measurement error is:

$$\Psi_{LS}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{X}_i, \Theta) = \begin{pmatrix} [\mathbf{Y}_i - (1, {\mathbf{Z}_i}^t, {\mathbf{X}_i}^t)\Theta_1]\mathbf{T}_i \\ (\frac{m-p}{m})\sigma^2 - [\mathbf{Y}_i - (1, {\mathbf{Z}_i}^t, {\mathbf{X}_i}^t)\Theta_1]^2 \end{pmatrix} \tag{4.6}$$

where $\mathbf{T}_i = (1, \mathbf{Z}_i, \mathbf{X}_i)^t$ and $p = dim(\Theta_1)$ and $\frac{m-p}{m}$ implements degrees of freedom correction for estimator of σ^2.

Now for $b = 1, ..., B$ generate random variables $\mathbf{Q}_{b,i}$ that are independent normal random vectors with mean zero and covariance matrix Σ_{uu}. Consider the complex-valued random variate $\tilde{\mathbf{W}}_{b,i} = \mathbf{W}_i + \mathbf{i}\mathbf{Q}_{b,i}$

The Monte Carlo corrected score (MCCS) is obtained in three steps:

- Replace X_i with $\tilde{\mathbf{W}}_{b,i}$ in a score function that is unbiased in the absence of measurement error. For linear least squares regression this is given by Equation (4.6).

- Take the real part, Re(.), of the resulting expression to eliminate the imaginary part.

- Average over multiple sets of pseudorandom vectors, $b = 1, ..., B$.

In linear regression, after these steps we obtain,

$$\tilde{\Psi}_{MCCS,B}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i, \Theta) = B^{-1} \sum_{b=1}^{B} Re[\Psi_{LS}(\mathbf{Y}_i, \mathbf{Z}_i, \tilde{W}_{b,i}, \Theta]$$

It can be shown that for all i and B,

$$E\{\tilde{\Psi}_{MCCS,B}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i, \Theta)/\mathbf{Z}_i, \mathbf{X}_i\} = \Psi_{LS}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{X}_i, \Theta)$$

and, hence, ignoring the degrees-of-freedom-correction we have:

$$E\{\tilde{\Psi}_{MCCS,B}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i, \Theta)\} = \mathbf{0}$$

The above equation shows that the corrected score $\Psi_{MCCS,B}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i, \Theta)$ is an unbiased estimator of the score Ψ_{LS} that would have been used if the measurement error was not present.

Carroll *et al.* (2006) showed that from general M-estimator theory and under regularity conditions the estimating equations given by:

$$\sum_{i=1}^{m} \tilde{\Psi}_{MCCS,B}(\mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i, \Theta) = \mathbf{0} \tag{4.7}$$

admit a consistent and asymptotically normal sequence of solutions and hence can be solved to obtain estimates of Θ_1 and σ^2 defined previously.

The estimates obtained using this Monte-Carlo corrected scores in this model are:

$$\widehat{\Theta}_1 = (\widehat{M}_{1zw,1zw} - \tilde{\Omega})^{-1}\widehat{M}_{y,1zw}, \hat{\sigma}^2 = (m-p)^{-1} \sum_{i=1}^{m} \{(\mathbf{Y}_i - \widehat{Y}_i)^2 - \hat{\beta}_x^T \hat{\Sigma}_{uu} \hat{\beta}_x\},$$

where,

$$\hat{M}_{1zw,1zw} = m^{-1}\sum_{i=1}^{m}\begin{pmatrix} 1 & \mathbf{Z}_i^t & \mathbf{W}_i^t \\ \mathbf{Z}_i & \mathbf{Z}_i\mathbf{Z}_i^t & \mathbf{Z}_i\mathbf{W}_i^t \\ \mathbf{W}_i & \mathbf{W}_i\mathbf{Z}_i^t & \mathbf{W}_i\mathbf{W}_i^t \end{pmatrix},$$

$$\widetilde{\Omega} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\Sigma}_{uu} \end{pmatrix}, \hat{\Sigma}_{uu} = (m^{-1}\sum_{i=1}^{m}\widehat{M}_{u,i}),$$

$$\widehat{M}_{y,1zw} = m^{-1}\sum_{i=1}^{m}\mathbf{Y}_i\begin{pmatrix} 1 \\ \mathbf{Z}_i \\ \mathbf{W}_i \end{pmatrix}, \widehat{Y}_i = (1, \mathbf{Z}_i^t, \mathbf{W}_i^t)\widehat{\Theta}_1,$$

$$\hat{\Sigma}_{uu} = (mB)^{-1}\sum_{i=1}^{m}\sum_{b=1}^{B}U_{b,i}U_{b,i}^t.$$

Replacing Σ_{uu} with $\hat{\Sigma}_{uu}$ yields, apart from the degrees of freedom corrections on relevant covariance matrices, the usual linear models, method of moments correction for measurement error bias (Fuller, 1987). Practically, this substitution can be avoided by taking B large, because $\hat{\Sigma}_{uu}$ converges to Σ_{uu} as $B \to \infty$. Usually, in practice, B does not need to be very large because the randomness that occurs in the construction of the Monte Carlo corrected scores is double-averaged over m and B.

MCCS and SIMEX: There is clearly a connection between SIMEX and Monte Carlo corrected scores since both involve the generating of pseudovariables via simulation. SIMEX adds measurement error multiple times, computes the estimator over each generated data set, and then extrapolates back to the case of no measurement error. The Monte Carlo corrected-score method is similar in spirit to SIMEX buts involves the reordering of the steps that constitute the SIMEX algorithm, namely, generation of pseudorandom and remeasured data sets, calculation of average estimates from those remeasured data sets and finally extrapolation to the case of no measurement error. Monte Carlo corrected-score method starts with complex-valued, pseudo-random data sets. Novick and Stefanski (2002) summarized the difference between the two as follows: rather than calculate an estimate from each complex pseudodata set and averaging, pseudodata *estimating equations* are averaged, the imaginary part is removed, and then the averaged equations are solved, resulting in a single estimate.

4.2. Corrected Score Estimators for Fay-Herriot Model

In Section 2.5, we considered a general linear model for estimation of parameters of small area given large scale survey data covering small areas and area level auxiliary information from the census or administrative records. In this section, we shall obtain bias corrected estimators for parameters of the Fay-Herriot model using the corrected score approach.

Consider the Fay-Herriot model given by:

$$y_i = X_i^t \beta + v_i + e_i, \tag{4.8}$$

for $i = 1, 2, ..., m$, X_i's are fixed and $v_i \sim Normal(0, \sigma_v^2)$, $e_i \sim Normal(0, \psi_i)$ and v_i's and e_i's are independently distributed. ψ_i's are known but σ_v^2 is unknown. In matrix notation the Fay-Herriot model can be written as:

$$\underset{\sim}{\mathrm{y}}_{m\times 1} = \underset{\sim}{\mathrm{X}}\beta + \underset{\sim}{\mathrm{v}} + \underset{\sim}{\mathrm{e}}, \tag{4.9}$$

such that independent $\underset{\sim}{v}$ and $\underset{\sim}{e}$ are distributed as $N_m(0, \sigma_v^2 I)$ and $Normal_m(0, \Sigma)$ respectively, where

$$\Sigma = Diag(\psi_1, \psi_2, ..., \psi_m). \tag{4.10}$$

Remark: The model considered in the Section 2.5 is more flexible in the following sense. Instead of considering known $\Sigma = Diag(\psi_1, \psi_2, ..., \psi_m)$, consider that sampling errors have variance-covariance matrix $D^* = Diag(\sigma_v^2\psi_1, \sigma_v^2\psi_2, ..., \sigma_v^2\psi_m) = \sigma_v^2\Sigma$ where σ_v^2 is the unknown variance considered for model error. Since both v_i and e_i are errors of the same area assuming such a variance relationship seems to be reasonable. Then, considering the transformation $\Sigma^{-1/2}y$, $\Sigma^{-1/2}X$, $\Sigma^{-1/2}v$, $\Sigma^{-1/2}e$, will reduce this Fay-Herriot model in the matrix notation with the above error structure to the model of Section 2.5. Since Σ is positive definite a unique $\Sigma^{-1/2}$ will exist.

The marginal distribution of y is multivariate normal $N_m(X\beta, \sigma_v^2 + \Sigma)$ where Σ is given in Equation (4.10). The marginal distribution of $\underset{\sim}{v}$ is $Normal_m(0, \sigma_v^2 I)$ and that of e is $N_m(0, \Sigma)$. The conditional distribution of $\underset{\sim}{y}$ given v is $N_m(X\beta + v, \Sigma)$. Therefore, the likelihood of sample observations is given by:

$$\begin{aligned} L(\beta; \underset{\sim}{v}, X, Y) = f(v).f(Y|v). = \mathbb{C} \times exp\Bigg[&\frac{-1}{2}(y - X\beta - v)^t(y - X\beta - v) \\ &- \frac{1}{2\sigma_v^2}v^t v\Bigg], \end{aligned} \tag{4.11}$$

where

$$\mathbb{C}^{-1} = [(2\pi)^m |\Sigma|^{1/2} \sigma_v^m]. \tag{4.12}$$

Thus the log-likelihood is obtained as:

$$l(\beta; v, X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}(y - X\beta - v)^t \Sigma^{-1}(y - X\beta - v) - \frac{1}{2\sigma_v^2}v^t v, \tag{4.13}$$

where $K(\sigma_v^2, \Sigma)$ is given by:

$$K(\sigma_v^2, \Sigma) = -m \times ln(2\pi\sigma_v) - \frac{1}{2}log|\Sigma|. \tag{4.14}$$

With some algebra, Equation (4.13) simplifies to:

$$\begin{aligned} l(\beta; v, X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}(y - X\beta)^t\Sigma^{-1}(y - X\beta) \\ + v^t\Sigma^{-1}(y - X\beta) - \frac{v^t\Sigma^{-1}v}{2} - \frac{1}{2\sigma_v^2}v^tv. \end{aligned} \tag{4.15}$$

Differentiating $l(\beta; v, X, Y)$ partially with respect to to β and equating the expression thus obtained to zero, we get,

$$\left(\Sigma^{-1} + \frac{I}{\sigma_v^2}\right) v = \Sigma^{-1}(y - X\beta),$$

or,

$$\begin{aligned} \hat{v} &= \left(\Sigma^{-1} + \frac{I}{\sigma_v^2}\right)^{-1} \Sigma^{-1}(y - X\beta) \\ &= \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} (y - X\beta) \end{aligned} \tag{4.16}$$

Substituting this value of $\hat{v}$ in Equation (4.15), we have, after some algebra,

$$\begin{aligned} l(\beta; v, X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}(y - X\beta)^t \Bigg\{ \Sigma^{-1} - 2\left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} \Sigma^{-1} \\ \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} \Sigma^{-1} \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} \Bigg\} (y - X\beta) \\ - \frac{1}{2\sigma_v^2}(y - X\beta)^t \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} (y - X\beta). \end{aligned} \tag{4.17}$$

Let,

$$A = \left\{ I + \frac{\Sigma}{\sigma_v^2} \right\}^{-1}. \tag{4.18}$$

The matrix Σ is symmetric and so is A. Using this fact the expression in Equation (4.17) can be simplified after extensive algebra to,

$$\begin{aligned} l(\beta; v, X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}(y - X\beta)^t D(y - X\beta) \\ - \frac{1}{2\sigma_v^2}(y - X\beta)^t H(y - X\beta), \end{aligned} \tag{4.19}$$

where,

$$H = A^2, \tag{4.20}$$

and,

$$D = (1 - A)^2 \Sigma^{-1}. \tag{4.21}$$

The likelihood for β at Equation (4.19) can finally be written as,

$$l(\beta; X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}(y - X\beta)^t \mathbb{P}(y - X\beta), \tag{4.22}$$

where,

$$\mathbb{P} = D + \frac{H}{\sigma_v^2}. \tag{4.23}$$

Note that $\mathbb{P}$ is also symmetric because D and H are so. The expression in Equation (4.22) can also be written as,

$$l(\beta; X, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}y^t \mathbb{P} Y + \beta^t X^t \mathbb{P} y - \frac{1}{2}\beta^t (X^t \mathbb{P} X)\beta. \tag{4.24}$$

Before proceeding any further, we obtained simple expressions for H, D and $\mathbb{P}$. Given $\Sigma = Diag(\psi_1, \psi_2, ..., \psi_m)$ the following results can easily be checked.

$$\begin{aligned} A &= \left(I + \frac{\Sigma}{\sigma_v^2}\right)^{-1} \\ &= Diag\left\{\frac{\sigma_v^2}{\sigma_v^2 + \psi_1}, \frac{\sigma_v^2}{\sigma_v^2 + \psi_2}, ..., \frac{\sigma_v^2}{\sigma_v^2 + \psi_m}\right\} \end{aligned} \tag{4.25}$$

Hence,

$$H = A^2 = Diag\left\{\frac{\sigma_v^4}{(\sigma_v^2+\psi_1)^2}, \frac{(\sigma_v^4}{\sigma_v^2+\psi_2)^2}, ..., \frac{\sigma_v^2}{(\sigma_v^2+\psi_m)^2}\right\} \tag{4.26}$$

The matrix $D = (1-A)^2\Sigma^{-1}$ is given by:

$$D = Diag\left\{\frac{\psi_1}{(\sigma_v^2+\psi_1)^2}, \frac{\psi_2}{(\sigma_v^2+\psi_2)^2}, ..., \frac{\psi_m}{(\sigma_v^2+\psi_m)^2}\right\}, \tag{4.27}$$

Finally, $\mathbb{P} = D + \frac{H}{\sigma_v^2}$ is given by,

$$\mathbb{P} = Diag\left\{\frac{1}{(\sigma_v^2+\psi_1)}, \frac{1}{(\sigma_v^2+\psi_2)}, ..., \frac{1}{(\sigma_v^2+\psi_m)}\right\} \tag{4.28}$$

and also,

$$tr(\mathbb{P}) = \sum_{i=1}^{m} \frac{1}{\sigma_v^2+\psi_i}. \tag{4.29}$$

Differentiating Equation (4.24) with respect β to obtain score function for β as follows:

$$\Psi(\beta; X, y) = \frac{\partial}{\partial\beta} l(\beta; X, Y) = X^t\mathbb{P}y - X^t\mathbb{P}X\beta. \tag{4.30}$$

Since we have observed W instead of set of covariates X where W is contaminated with non-differential additive classical error, we obtain $l(\beta; X, Y)$ and $\Psi(\beta; X, Y)$ by replacing X by W in Equations (4.25) and (4.30) respectively.

Thus we have,

$$l(\beta; W, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}y^t\mathbb{P}y + \beta^t W^t\mathbb{P}y - \frac{1}{2}\beta^t(W^t\mathbb{P}W)\beta, \tag{4.31}$$

and,

$$\Psi(\beta; W, y) = \frac{\partial}{\partial\beta} l(\beta; W, Y) = W^t\mathbb{P}y - W^t\mathbb{P}W\beta. \tag{4.32}$$

In the Section 2.5 we have already shown that $E^*(W) = X$ and $E^*W^t\mathbb{P}W = X^t\mathbb{P}X +$

$tr(\mathbb{P})\Lambda$. Thus, taking conditional expectation E^* with respect to W given v, y and Λ we have:

$$\begin{aligned} E^*\{l(\beta, W, Y)\} &= K(\sigma_v^2, \Sigma) - \frac{1}{2}y^t\mathbb{P}y + \beta^t X^t\mathbb{P}y - \frac{1}{2}\beta^t(X^t\mathbb{P}X + tr(\mathbb{P})\Lambda)\beta \\ &= K(\sigma_v^2, \Sigma) - \frac{1}{2}y^t\mathbb{P}y + \beta^t X^t\mathbb{P}y - \frac{1}{2}\beta^t X^t\mathbb{P}X\beta - \frac{1}{2}tr(\mathbb{P})\beta^t\Lambda\beta. \end{aligned} \tag{4.33}$$

and,

$$E^*\{\Psi(\beta; W, y)\} = X^t\mathbb{P}y - W^t\mathbb{P}W\beta - tr(\mathbb{P})\Lambda\beta. \tag{4.34}$$

The corrected log-likelihood and score functions must satisfy the conditions:

$$E^*\left\{\frac{\partial}{\partial\beta}l^*(\beta, W, y)\right\} = \frac{\partial}{\partial\beta}l(\beta, X, y), \tag{4.35}$$

and,

$$E^*\{\Psi^*(\beta, W, y)\} = \Psi(\beta, X, y). \tag{4.36}$$

From Equation (4.33) we obtain,

$$l^*(\beta, W, Y) = K(\sigma_v^2, \Sigma) - \frac{1}{2}y^t\mathbb{P}y + \beta^t W^t\mathbb{P}y - \frac{1}{2}\beta^t W^t\mathbb{P}W\beta + \frac{1}{2}tr(\mathbb{P})\beta^t\Lambda\beta, \tag{4.37}$$

which will satisfy both the conditions. Thus $l^*(\beta, W, Y)$ is a corrected log-likelihood for our model and the corrected score function can similarly be obtained from Equation (4.34) as follows:

$$\Psi^*(\beta; W, Y) = W^t\mathbb{P}y - W^t\mathbb{P}W\beta + tr(\mathbb{P})\Lambda\beta. \tag{4.38}$$

In order to obtain consistent estimator from Equation (4.38) we must check that $\Psi^*(\beta; W, Y)$ is an unbiased score function. To do this we will take the global expectation $E = E^+E^*$ of Equation (4.38) as follows,

$$\begin{aligned}
E\{\Psi^*(\beta; W, y)\} &= E^+[E^*\Psi^*(\beta; W, y)] \\
&= E^+[(E^*W)\mathbb{P}y - E^*(W^t\mathbb{P}W)\beta + tr(\mathbb{P})\Lambda\beta] \\
&= E^+\{X^t\mathbb{P}y - [X^t\mathbb{P}X + tr(\mathbb{P})\Lambda]\beta + tr(\mathbb{P})\Lambda\beta\} \\
&= E^+\{X^t\mathbb{P}y - X^t\mathbb{P}X\beta\} \\
&= X^t\mathbb{P}X\beta - X^t\mathbb{P}X\beta = 0.
\end{aligned}$$

Thus $\Psi^*(\beta; W, Y)$ is an unbiased corrected score function and will provide fully consistent estimator for β. Thus equating $\Psi^*(\beta; W, Y)$ to zero we have,

$$W^t\mathbb{P}y = \{W^t\mathbb{P}W - tr(\mathbb{P})\Lambda\}\beta.$$

Hence, the corrected score estimator in the Fay-Herriot model will be,

$$\hat{\beta}_{FHCS} = \{W^t\mathbb{P}W - tr(\mathbb{P})\Lambda\}^{-1}W^t\mathbb{P}y. \tag{4.39}$$

The condition,

$$E^*\left\{\frac{\partial}{\partial v} l^*(\beta; v, W, y)\right\} = \frac{\partial}{\partial v} l(\beta; v, X, y),$$

could also be easily verified for the corrected log-likelihood function given by:

$$\begin{aligned}
l^*(\beta; v, W, y) = K(\sigma_v^2, \Sigma) &- \frac{1}{2}(y - W\beta - v)^t\Sigma^{-1}(y - W\beta - v) \\
&- \frac{1}{2\sigma_v^2}v^tv + \frac{1}{2}tr(\mathbb{P})\beta^t\Lambda\beta.
\end{aligned} \tag{4.40}$$

The corrected score function shall be given by,

$$\begin{aligned}
\Psi^*(\beta; v, W, y) &= \frac{\partial}{\partial v} l^*(\beta; v, W, Y) \\
&= \Sigma^{-1}(Y - W\beta) - \Sigma^{-1}v - \frac{v}{\sigma_v^2}.
\end{aligned} \tag{4.41}$$

Equating $\Psi^*(\beta; v, W, y)$ to zero gives,

$$\left\{\Sigma^{-1} + \frac{I}{\sigma_v^2}\right\}v = \Sigma^{-1}(y - W\widehat{\beta}_{FHCS})$$

or, the corrected score estimator of random effect v in the Fay-Herriot model is:

$$\begin{aligned}\widehat{v}_{FHCS} &= \left\{\Sigma^{-1} + \frac{I}{\sigma_v^2}\right\}^{-1}\Sigma^{-1}(y - W\widehat{\beta}_{FHCS}) \\ &= \left\{I + \frac{\Sigma^{-1}}{\sigma_v^2}\right\}^{-1}(y - W\widehat{\beta}_{FHCS})\end{aligned} \tag{4.42}$$

Now we shall simplify expressions in order to obtain estimate for i^{th} small area. $\widehat{v}_{FHCS}$ in Equation (4.42) can be simplified using the expression for $A = \left\{I + \frac{\Sigma^{-1}}{\sigma_v^2}\right\}^{-1}$ given at Equation (4.25). Thus,

$$\widehat{v}_{FHCS} = Diag\left\{\frac{\sigma_v^2}{\sigma_v^2 + \psi_1}, \frac{\sigma_v^2}{\sigma_v^2 + \psi_2}, \ldots.., \frac{\sigma_v^2}{\sigma_v^2 + \psi_m}\right\} \times \begin{pmatrix} y_1 - W_1^t\widehat{\beta}_{FHCS} \\ y_2 - W_2^t\widehat{\beta}_{FHCS} \\ \vdots \\ \vdots \\ y_m - W_m^t\widehat{\beta}_{FHCS} \end{pmatrix}$$

Thus, we have,

$$\widehat{v}_{iFHCS} = \frac{\sigma_v^2}{\sigma_v^2 + \psi_i}(y_i - W_i^t\hat{\beta}_{FHCS}). \tag{4.43}$$

Moreover, using expression for $\mathbb{P}$ in Equation (4.28) we obtain,

$$W^t\mathbb{P}W - tr(\mathbb{P})\Lambda = \left\{\sum_{i=1}^{m}\frac{W_iW_i^t}{\sigma_v^2 + \psi_i} - tr(\mathbb{P})\Lambda\right\}, \tag{4.44}$$

and,

$$W^t\mathbb{P}y = \sum_{i=1}^{m} W_iy_i. \tag{4.45}$$

Substituting the values from Equations (4.44) and (4.45) in to Equation (4.39) we obtain the simplified value of $\widehat{\beta}_{iFHCS}$ as,

$$\widehat{\beta}_{iFHCS} = \left\{ \sum_{i=1}^{m} \frac{W_i W_i^t}{\sigma_v^2 + \psi_i} - tr(\mathbb{P})\Lambda \right\}^{-1} \sum_{i=1}^{m} W_i y_i. \tag{4.46}$$

The estimate for the parameter in the i^{th} area is:

$$\begin{aligned} \hat{\theta}_i &= W_i^t \widehat{\beta}_{iFHCS} + \widehat{v}_{iFHCS} \\ &= W_i^t \widehat{\beta}_{iFHCS} + \frac{\sigma_v^2}{\sigma_v^2 + \psi_i}(y_i - W_i^t \widehat{\beta}_{FHCS}) \\ &= \frac{\sigma_v^2}{\sigma_v^2 + \psi_i} y_i + \frac{\psi_i}{\sigma_v^2 + \psi_i} W_i^t \widehat{\beta}_{iFHCS} \\ &= \varphi_i y_i + (1 - \varphi_i) W_i^t \widehat{\beta}_{iFHCS}. \end{aligned} \tag{4.47}$$

where the weights φ_i are given by,

$$\varphi_i = \frac{\sigma_v^2}{\sigma_v^2 + \psi_i}. \tag{4.48}$$

Thus, $\hat{\theta}_i$ is the weighted average of direct survey design based unbiased estimator y_i and synthetic estimator $W_i^t \widehat{\beta}_{iFHCS}$ which borrows strength from other areas through auxiliary information.

Note that when there is no measurement error, that is, $\Lambda = 0$ and $W_i^T = X_i^t$ and the estimator reduces to conventional BLUP estimator.

4.2.1. Estimation of Variance Components

While estimating β and predicting v we assumed that σ_v^2, Λ are known and in practice they are replaced by precise estimates. We shall follow the approach suggested by Harville (1977) and Schall (1991) for estimating variances associated with all variance components involved in the model. First, we shall briefly describe their algorithm and then provide a suitable modification for our model. The approach in the general situation is as follows.

Consider the model,

$$y = X\beta + Z_1 v_1 + Z_2 v_2 + ... + Z_c v_c + e, \tag{4.49}$$

where y is an $n \times 1$ vector of observations, X is a full column rank $n \times p$ matrix, β is an unknown $p \times 1$ vector of fixed effects Z_i, $(i = 1, .., c)$ are known $n \times q_i$ known matrices, v_i is a $q_i \times 1$ unknown vector of random effects and e is an unknown vector of random errors.

For $i \neq j$ and $i = 1, 2, ..., c$ the error structure of y is assumed to satisfy $E(v_i) = 0$, $E(e) = 0$, $Cov(v_i, v_j) = 0$, $Cov(v_i, e) = 0$. In addition,

$$Var(v_i) = \mathbb{D}_i = \sigma^2_{iv} I_{q_i},$$

where I_{q_i}, is an identity matrix of order q_i and,

$$Var(e) = \mathbb{G} = \sigma^2_{(c+1)v} I_{n \times n}.$$

Let $q = q_1 + q_2 + ... + q_c$. Let Z be the $n \times q$ partitioned matrix,

$$Z = [Z_1, ..., Z_c],$$

and let v be the partitioned vector defined by,

$$v^t = (v_1^t, v_2^t, ..., v_c^t).$$

The model in Equation (4.49) can then be written as,

$$y = X\beta + Zv + e, \tag{4.50}$$

with $Var(v) = \mathbb{D} = Diag(\mathbb{D}_1, \mathbb{D}_2, ..., \mathbb{D}_c)$. Note that the mean and variance of y can be written as $E(y) = X\beta$, and,

$$\begin{aligned} Var(y) = \mathbb{V} &= \mathbb{G} + Z\mathbb{D}Z^t \\ &= \sum_{i=1}^{c} \sigma^2_{iv} Z_i Z_i^t + \sigma^2_{(c+1)v} I_{n\times n}. \end{aligned}$$

The estimators $\widehat{\sigma}^2_{1v}$, $\widehat{\sigma}^2_{2v}$, and $\widehat{\sigma}^2_{(c+1)v}$ of σ^2_{1v}, σ^2_{2v}, and $\sigma^2_{(c+1)v}$ are obtained by a simple modification of the restricted maximum likelihood estimation method.

The algorithm runs as follows:

- Step 1: Given estimates $\widehat{\sigma}^2_{(c+1)v}$ and $\widehat{\sigma}^2_{1v}$, $\widehat{\sigma}^2_{2v}$,...,$\widehat{\sigma}^2_{cv}$ compute estimates β and $\hat{v}_1, \hat{v}_2, ..., \hat{v}_c$ as least squares solutions to the set of equations,

$$\mathbb{C}\begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} \mathbb{G}^{-1/2}X & \mathbb{G}^{-1/2}Z \\ 0 & \mathbb{D}^{-1/2} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} \mathbb{G}^{-1/2}y \\ 0 \end{pmatrix}, \tag{4.51}$$

  where $\mathbb{G} = \sigma^2_{(c+1)v} I_{n\times n}$ and $\mathbb{D} = Diag(\mathbb{D}_1, \mathbb{D}_2, ..., \mathbb{D}_c)$. The $\mathbb{D}_i = \sigma^2_{iv}$ are evaluated at the current estimates of the variance components. Note that the best linear unbiased estimators of β and v are given by Equations (4.43) and (4.41) respectively. (Henderson, 1963).

- Step 2: Let $\mathbb{T}$ be the matrix formed from the last q rows and last q columns of $[\mathbb{C}^T\mathbb{C}]^{-1}$ partitioned similar to $\mathbb{D}$:

$$\begin{pmatrix} \mathbb{T}_{11} & \mathbb{T}_{12} & ..... & \mathbb{T}_{1c} \\ : & : & : & : \\ : & : & : & : \\ \mathbb{T}_{c1} & \mathbb{T}_{c2} & ..... & \mathbb{T}_{cc} \end{pmatrix}, \tag{4.52}$$

and,

$$\mathbb{C}_i = \frac{tr(\mathbb{T}_{ii})}{\sigma^2_{iv}}. \tag{4.53}$$

The restricted maximum likelihood estimators σ^2_{1v}, σ^2_{2v},..,$\sigma^2_{(c+1)v}$ satisfy,

$$\sigma^2_{iv} = \frac{v_i^t v_i}{q - \mathbb{C}_i}, \tag{4.54}$$

for $i = 1, 2, ..., c$ and

$$\sigma^2_{(c+1)v} = \frac{\hat{e}_i^t \hat{e}_i}{n - p - q^*}, \tag{4.55}$$

where $q^* = \sum_{i=1}^{c}(q_i - \mathbb{C}_i)$ and,

$$\hat{e}_i = y_i - X\hat{\beta} - Z\hat{v}. \tag{4.56}$$

Here $\mathbb{C}_i = tr(\mathbb{T}_{ii})/\sigma^2_{iv}$ is evaluated at the current estimate of σ^2_{iv} and estimators from equations (4.53) and (4.54) are obtained by setting the derivatives of the restricted likelihood to zero. The algorithm is repeated till we obtain stable estimates of variance components.

In our case of the Fay-Herriot model, we shall consider the set of equations corrected for the measurement error instead of Equations (4.51). These equations, as described in the previous section, are obtained by equating to zero, the partial derivatives of corrected log-likelihoods $l^*(\beta, v, \mathbf{X}, \mathbf{Y})$ given by Equation (4.40). in the previous section, in the Fay-Herriot model with respect to β and v. The set of equations turn out to be:

$$\begin{pmatrix} W^t\Sigma^{-1}W - tr(\mathbb{P}).\Lambda & W^t\Sigma^{-1} \\ \Sigma^{-1}W & \Sigma^{-1} + \frac{1}{\sigma^2_v} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} W^t\Sigma^{-1}y \\ \Sigma^{-1}y \end{pmatrix}. \tag{4.57}$$

Note that $tr(\mathbb{P}) = \sum_{i=1}^{m} 1/(\sigma_v^2 + \psi_i)$ also depends on σ_v^2. The algorithm is repeated until we obtain stable estimates. After the convergence is obtained, the estimate of σ_v^2 is given as,

$$\widehat{\sigma}_v^2 = \frac{\hat{v}^t \hat{v}}{m - \mathbb{C}_i}. \tag{4.58}$$

Since we have assumed normality of errors, the least square and likelihood estimating equations will be the same. Alternatively, if we equate the partial derivative of $l^*(\beta, v, \mathbf{X}, \mathbf{Y})$ given by Equation (4.40) in the previous section, with respect to σ_v^2 we obtain,

$$\widehat{\sigma}_v^2 = \frac{\hat{v}^t \hat{v}}{m} - \frac{1}{m} \sum_{i=1}^{m} \frac{\widehat{\beta}^t \Lambda \widehat{\beta}}{\left\{ 1 + \frac{\psi_i}{\widehat{\sigma}_v^2} \right\}^2}. \tag{4.59}$$

We use an iterative approach to solve Equation (4.57), starting with the initial estimate of β where the sample estimate of $\hat{\sigma}_v^2$ is the same as the one used in EBLUP (Rao, 2003b). Finally, Λ is estimated using the method of moments. (Fuller, 1987; Carroll *et al.*, 1995). Note: In the GLMM considered in Section 2.5, the estimate of σ_i^2 using this method will be,

$$\widehat{\sigma}_i^2 = \frac{(y_i - W_i \hat{\beta} - \hat{v}_i)^t (y_i - W_i \hat{\beta} - \hat{v}_i)}{m - p - q^*}. \tag{4.60}$$

4.3. Properties

The following are some of the properties of corrected scores estimators. These properties are obtained for corrected score estimators in the more general linear mixed model case discussed in Section 2 of this chapter and can easily be extended to the Fay Herriot model. All technical details are included in Appendix B.

Result 1: Along the lines of Proposition 1 in Nakamura (1990), suppose that $l^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ and $l(\beta; \mathbf{X},\ \mathbf{Y})$ are differentiable and

$$\sum_{i=1}^{m} i^{-2} Var^*\{l^*(\beta; W_i, Y_i)\} < \infty,$$

Also, suppose that β is identifiable and Y_i's are mutually independent. Then the corrected score function $S^*(\beta; \mathbf{W},\ \mathbf{Y})$ has a root which is consistent with probability one as $m \to \infty$

Remark: This above result is for the purpose of ensuring the identifiability of the β. Fuller (1987, p. 103) describes the inconsistency of a maximum likelihood estimate for a linear model accounting for measurement errors when X_i's are fixed vectors; his likelihood treats X_i's as unknown parameters and therefore the number of parameters to be estimated increases with m. On the other hand, the corrected log likelihood in our case of the Fay-Herriot model, does not involve the X_i's and consequently the X_i's need not be estimated to obtain the root of the corrected score.

Result 2: If $\hat{\beta}$ and β are consistent roots of $S^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y}) = 0$ and $S(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y}) = 0$ and $S^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ and $S(\beta; \mathbf{V},\ \mathbf{X},\ \mathbf{Y})$ satisfy some regularity conditions then the conditional distribution of $(\hat{\beta} - \beta)$ given $\mathbf{X}$ and $\mathbf{Y}$ is asymptotic normal with mean zero and variance- covariance matrix:

$$I^+(\beta, \mathbf{X})^{-1} E^+[Var^*\{S^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})\}] I^+(\beta, \mathbf{X})^{-1},$$

as $m \to \infty$.

Remark : We shall omit the details of the proof of result 2, since we have assumed that $S^*(\beta; \mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ is totally differentiable and a sum of m independent contributions, so the result follows easily on Taylor series expansion of $S^*(\beta; \mathbf{V}, \mathbf{W}, \mathbf{Y})$ at $\beta = \beta_x$ and application of Cramer's technique of proof (Scholtz, 1981, p. 340-51; Cox and Hinkley, 1974, p. 294-6; Rao, 1973, p. 364-5).

Result 3: Assume that all derivatives with respect to likelihood exist, R^{-1} is positive definite (where $R^{-1} = I + (I - \Sigma^{-1})^{-1}$) and that the parameter β is identifiable. Assume that as $m \to \infty$, the following limits exist: $m^{-1}\mathbf{X}^T R^{-1}\mathbf{X}$, $m^{-1}\mathbf{X}^T I_m$, $m^{-1}(I + \Sigma^{-1})$, $m^{-1}\mathbf{X}^T R^{-2}\mathbf{X}$, $m^{-1}tr(R^{-1})$ and $m^{-1}tr(R^{-2})$, then:

$$\mathbf{W}^T R^{-1}\mathbf{W} = \mathbf{X}^T R^{-1}\mathbf{X} + tr(R^{-1})\Lambda + O_p(m^{1/2}).$$

Remark: The existence of all the limits mentioned above as $m \to \infty$ is ensured by Lee and Nelder (1996).

Result 4: In addition to all the conditions in Result 3, let us assume that all the regularity conditions for the existence of maximum likelihood estimates hold. Then the corrected scores estimator $\widehat{\beta}_{FHCS}$ derived for our Fay-Herriot model is asymptotically normally distributed. The asymptotic mean and variance of $\widehat{\beta}_{FHCS}$ are respectively given as β and:

$$\begin{aligned} avar(\widehat{\beta}_{FHCS}) &= (\mathbf{X}^T R^{-1}\mathbf{X})^{-1}(\sigma_v^2\Sigma + G)(\mathbf{X}^T R^{-2}\mathbf{X})(\mathbf{X}^T R^{-1}\mathbf{X})^{-1} \\ &\quad + (\mathbf{X}^T R^{-1}\mathbf{X})^{-1}\mathbf{J}(\mathbf{X}^T R^{-1}\mathbf{X})^{-1}, \end{aligned}$$

respectively where $\mathbf{J} = \{tr[R^{-2}(\sigma_v^2\Sigma + G)] + \beta^T(\mathbf{X}^T R^{-2}\mathbf{X})\beta\}\Lambda$.

Corollary: Let β be the true value of the regression coefficient then $\widehat{\beta}_{FHCS}$ is consistent in probability and $\hat{\beta}_{FHCS} - \beta = O_p(m^{-1/2})$.

Result 5: Let $\hat{V} = \mathbf{V}_{FHCS}(\beta)$. Then assuming that all the conditions and properties given in results 3 and 4 hold, $\mathbf{V}_{FHCS} - \mathbf{V} = O_p(m^{-1/2})$ and it is asymptotically normally distributed with asymptotic variance given by:

$$avar(\mathbf{V}_{FHCS} - \mathbf{V}) = \mathbf{J}_1^{-1}\mathbf{J}_2\{avar(\hat{\beta}_{FHCS})\}\mathbf{J}_2^T\mathbf{J}_2^{-1},$$

where $\mathbf{J}_1 = m^{-1}(I + \Sigma^{-1})$ and $\mathbf{J}_2 = m^{-1}(I^T\mathbf{X})$.

Using these corrected score estimators of regression coefficients β and random effects $\mathbf{V}$ the estimate of population total in i^{th} small area is calculated as $\hat{Y}_{iFHCS} = X_i^t\hat{\beta}_{FHCS} + \hat{V}_{iFHCS}$, and the Monte Carlo corrected score estimator as $\hat{Y}_{iMCCS} = X_i^t\hat{\beta}_{MCCS} + \hat{V}_{iMCCS}$, and its performance is studied via simulation in the next section and compared to the SIMEX, Lohr-Ybarra, direct and MCCS estimators in a variety of scenarios. The estimates of nuisance parameters like the random effect and measurement error variances will be discussed in a later section. The Monte Carlo corrected score estimators are obtained using the procedure described in the Section 4.1.1. of this chapter, the difference in the Fay-Herriot model being that there are no covariates measured without error, but there are random effects present. We shall study the theoretical properties of the corrected scores estimator in the Fay-Herriot model and the theoretical details for Monte Carlo corrected score estimators will be omitted and can be found in Novick and Stefanski's paper (2002).

4.4. Simulation Study

A simulation study was conducted after the calculation of corrected scores estimators of β and random effects $\mathbf{V}$ and consequently the estimate of population total in i^{th} small area for bias correction in the above model. This simulation study is set up on the same lines as Chapter III, for the purpose of meaningful comparison. We generated $X_i \sim N(4, 9)$, $\psi_i \sim$ Gamma(4, 2). For each iteration we generated $Y_i = 1 + 4x_i + v_i$, $y_i = Y_i + e_i$ and $\widehat{x}_i = x_i + u_i$, where v_i, e_i and u_i are independent normal variables with mean 0 and variance σ_v^2, ψ_i and c_i respectively. Consider 3 factors (Lohr and Ybarra, 2008) : Factor 1: $\sigma_v^2 = 2$, 3 or 4; Factor 2: $c_i \in \{0, d\}$ for d= {2,3, or 4}; Factor 3: m= 20, 50 or 100. The number of iterations for each combination were 10000. As in Chapter II, three different scenarios w.r.t. X_i were considered, i.e., ALL of them being measured with error (k=100), some (specified percentage k) measured with error and NONE ($k = 0$) of them measured with error. SIMEX and MCCS estimates were obtained after generating pseudo-variables as described in previous chapters and sections.

We found empirical MSE's, for each area i, for the direct, Fay-Herriot estimator that ignores measurement error and treats $\widehat{X}_i$ as the actual observed X_i, Lohr-Ybarra, SIMEX, corrected scores and MCCS estimators. These empirical MSE's are calculated as, $\Sigma_{l=1}^{10000}\Sigma_i(\widehat{Y}_{i(l)} - Y_{i(l)})^2/10000$ where $Y_{i(l)}$ and $\widehat{Y}_{i(l)}$ are the true and predicted values of $X_i^T\beta + v_i$ in l^{th} iteration. For the Lohr-Ybarra and SIMEX estimators, the estimated weights $\widehat{\gamma}_{iv}$ described in Chapter III, the consistent estimator $\widehat{\sigma}_v^2$ provided by result 4 in Section 3 of Chapter III is used and while calculating $\widehat{Y}_{iSIMEX}$ the SIMEX estimator of β is substituted in place of the weighted least squares estimator of β. The results for $m = 100$, $C_i = 3$, $\sigma_v^2 = 4$ are displayed in Table 11.

Table 11. Empirical mean squared error for the six estimators, y_i, $\widetilde{Y}_{iS}$, $\widehat{Y}_{iME}$, $\widehat{Y}_{iSIMEX}$, $\widehat{Y}_{iFHCS}$, $\widehat{Y}_{iMCCS}$ when the number of small areas is 100, measurement error variance $C_i = 3$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	C_i	y_i	$\widetilde{Y}_{iS}$	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	0	8.3	3.8	3.7	3.8	3.7	3.7
20	3	10.2	7.3	6.4	4.2	4.0	4.1
50	3	8.9	5.1	6.4	3.2	3.7	3.6
80	3	10.8	7.7	7.4	5.8	5.9	5.7
100	3	10.9	6.5	6.6	3.7	3.8	3.6

Table 12. Absolute value of the bias for the six estimators, y_i, $\widetilde{Y}_{iS}$, $\widehat{Y}_{iME}$, $\widehat{Y}_{iSIMEX}$, $\widehat{Y}_{iFHCS}$, $\widehat{Y}_{iMCCS}$ when the number of small areas is 100, measurement error variance $C_i = 2$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	y_i	$\widetilde{Y}_{iS}$	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	1.31	0.22	0.21	0.23	0.22	0.23
20	1.72	0.41	0.65	0.28	0.29	0.30
50	1.76	0.72	0.46	0.38	0.39	0.38
80	1.64	0.82	0.54	0.28	0.29	0.27
100	1.72	0.85	0.63	0.29	0.28	0.27

In the Tables 11 and 12, six estimators are compared. y_i is the direct estimator, $\widetilde{Y}_{iS}$ that blindly uses the Fay-Herriot estimator $\widetilde{Y}_{iS} = \widehat{\gamma}_{iv} y_i + (1 - \widehat{\gamma}_{iv}) \widehat{X}_i^T \beta$ in which $\widehat{\gamma}_{iv}$ and $\widehat{\beta}$ are calculated assuming $\widehat{X}_i$'s are the true values, $\widehat{Y}_{iME}$ is the Lohr-Ybarra estimator, $\widehat{Y}_{iSIMEX}$ is our SIMEX estimator described in detail in Chapter II. In Table 12, absolute value of the bias is reported for the estimators when m=100, C_i= 2 ; $\sigma_v^2 = 4$ and k as before. Once again we arrive at the conclusion that the direct estimator y_i performs the worst almost for all cases, i.e., these estimates have very large average empirical mean squared errors and hence one should refrain from using them if auxiliary information is available. The SIMEX estimate and the corrected score estimate both perform much better than the Lohr-Ybarra estimator. The Monte Carlo corrected score estimator and SIMEX estimator seem to have mean squared errors that are very close in value, which is expected since their nature is quite similar as discussed in a previous section. But in the worst case scenario, that is when $k = 100$ and ALL the areas have auxiliary information measured with error, the Monte Carlo corrected score estimators are marginally better than the SIMEX estimators.is measured with error. As the proportion of areas having auxiliary information measured with error increases, the empirical mean squared error of the SIMEX, corrected score and MCCS estimator is seen to be lower than that of the estimators that have been suggested before. Again, the fallacy of ignoring measurement error in covariates and using the Fay-Herriot estimator even when the exact X_i have not been observed, is brought to light by the increase in the mean squared error and bias of $\widetilde{Y}_{iS}$ as k increases.

Now we shall compare the jackknife estimates of mean squared errors of the Lohr-Ybarra, SIMEX, corrected score and MCCS estimators and assess their performance.

Table 13. Jackknife estimates of the mean squared errors of the Lohr-Ybarra estimator $\widehat{Y}_{iME}$, the SIMEX estimator $\widehat{Y}_{iSIMEX}$, the corrected scores estimator $\widehat{Y}_{iFHCS}$ and the Monte Carlo corrected score estimator $\widehat{Y}_{iMCCS}$ when the number of small areas is 50, measurement error variance $C_i = 4$ and $\sigma_v^2 = 4$. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	3.5	3.4	3.4	3.5
20	4.6	3.9	3.8	3.8
50	5.9	3.7	3.8	3.7
80	5.6	3.5	3.5	3.4
100	7.3	3.4	3.3	3.1

Table 14. Jackknife estimates of the mean squared error of the Lohr-Ybarra estimator $\widehat{Y}_{iME}$, the SIMEX estimator $\widehat{Y}_{iSIMEX}$, the corrected scores estimator $\widehat{Y}_{iFHCS}$ and the Monte Carlo corrected score estimator $\widehat{Y}_{iMCCS}$ when the number of small areas is relatively small, i.e., 20, measurement error variance $C_i = 2$ and $\sigma_v^2 = 3$. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	4.4	4.3	4.3	4.4
20	5.8	5.6	5.5	5.6
50	6.1	4.9	4.8	4.8
80	6.8	4.5	4.4	4.3
100	7.2	4.1	4.2	4.0

Tables 13 and 14 of the jackknife estimates of the mean squared errors also point towards the fact that the SIMEX, corrected score and Monte Carlo corrected score estimators perform much better than the Lohr-Ybarra estimators and the Monte-Carlo corrected score estimator preforms very marginally better than the rest when almost ALL small areas have auxiliary information measured with error. This superiority is unaffected by the low number of small areas, 20, in Table 14.

4.4.1. Studying Departure from Normality of Measurement Error

As done in Chapter II, a simulation study was also conducted by generating the measurement error U_i from a heavy-tailed distribution like t-distribution with 10 and 5 degrees of freedom and skew normal distribution with skewness parameter equal to -1 and 2 . The absolute value of the bias was calculated for the Lohr-YBarra, SIMEX, corrected score and MCCS estimators.

It can be clearly seen that the both the Lohr-Ybarra estimator and corrected score estimator are not very robust to departures from normality. This was expected since we assumed the normality of all the errors in the Fay-Herriot model while using the log-likelihood equation to derive the corrected score estimator. The Monte Carlo corrected score estimator, however, steers clear of distributional assumptions and hence is more generally applicable. Therefore, we can see from the results in Tables 15, 16, 17 and 18 that the SIMEX estimator and MCCS estimators are both very robust to departure from normality of measurement error, the MCCS estimator being very marginally superior to the SIMEX estimator when the the proportion k is 100. In any case, all our three estimators are far superior to the Lohr-Ybarra estimator which seems to perform very poorly once normality of measurement errors does not exist.

Table 15. Absolute value of bias for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$, the SIMEX estimator $\widehat{Y}_{iSIMEX}$, the corrected scores estimator $\widehat{Y}_{iFHCS}$ and the Monte Carlo corrected score estimator $\widehat{Y}_{iMCCS}$ when the number of small areas is 50, $\sigma_v^2 = 3$ and the U_i are generated from a t-distribution with 10 degrees of freedom. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	1.39	1.28	1.30	1.27
20	1.45	1.33	1.40	1.32
50	1.56	0.72	0.82	0.70
80	1.57	0.70	0.88	0.69
100	1.93	0.69	0.87	0.68

Table 16. Empirical mean squared errors for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$, SIMEX estimator $\widehat{Y}_{iSIMEX}$, the corrected scores estimator $\widehat{Y}_{iFHCS}$ and the Monte Carlo corrected score estimator $\widehat{Y}_{iMCCS}$ when the number of small areas is 50, $\sigma_v^2 = 4$ and the U_i are generated from a t-distribution with 5 degrees of freedom. k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	5.87	5.86	5.88	5.87
20	5.34	5.21	5.32	5.10
50	6.68	5.20	6.45	5.21
80	6.73	4.79	6.46	4.78
100	6.74	4.72	6.45	4.71

Table 17. Absolute value of bias for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$, the SIMEX estimator $\widehat{Y}_{iSIMEX}$, the corrected scores estimator $\widehat{Y}_{iFHCS}$ and the Monte Carlo corrected score estimator $\widehat{Y}_{iMCCS}$ when the number of small areas is 50, $\sigma_v^2 = 2$ and the U_i are generated from a skew-normal distribution with location parameter 0, scale parameter 3 and skewness parameter -1.5 (left-skewed distribution). k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	1.58	1.33	1.57	1.34
20	2.42	1.75	1.88	1.76
50	3.10	1.67	1.98	1.67
80	2.97	1.09	2.01	1.08
100	3.43	0.87	1.79	0.85

Tables 17 and 18 exhibit results from generation of skew normal measurement errors. The package ‘sn’ in R was used for this.

Table 18. Empirical mean squared errors for the Lohr-Ybarra estimator, $\widehat{Y}_{iME}$ and the SIMEX estimator $\widehat{Y}_{iSIMEX}$ when the number of small areas is 100, $\sigma_v^2 = 2$ and the U_i are generated from a skew-normal distribution with location parameter 0, scale parameter 3 and skewness parameter 2 (right-skewed distribution). k is the percentage of areas having auxiliary information measured with error.

k	$\widehat{Y}_{iME}$	$\widehat{Y}_{iSIMEX}$	$\widehat{Y}_{iFHCS}$	$\widehat{Y}_{iMCCS}$
0	6.04	6.05	6.37	6.04
20	5.85	5.35	5.97	5.33
50	5.78	5.34	5.64	5.32
80	5.79	4.98	5.50	4.96
100	6.28	4.87	5.99	4.86

4.5. Data Example

We apply the corrected scores and the Monte Carlo corrected scores method to the same data set used in Chapter III, namely the one consisting of body mass index values for different demographic subgroups from the 2003-2004 National Center for Health Statistics, U.S. National Health and Nutrition Examination Survey (NHANES) using the 2004 National Center for Health Statistics, U.S. National Health Interview Survey as auxiliary information. The jackknife estimates of the mean squared errors of the corrected score estimator were calculated in R version 2.1.1 and for the MCCS estimator using both Matlab and R.

Recall from Chapter III that the SIMEX method produced the estimate of the regression coefficient as 0.94 and intercept as 0.16. The corresponding estimate of σ_v^2 using the SIMEX method was 0.52. The corrected score estimates for the regression coefficient and intercept were 0.86 and 0.17 respectively. $\hat{\sigma}_v^2$ using the corrected score approach was 0.65. The Monte Carlo corrected score estimates for the regression coefficient and intercept were 0.93 and 0.16 respectively. $\hat{\sigma}_v^2$ using the corrected score approach was 0.55.

In Figure 4, we plot the jackknife estimates of the mean squared error for the corrected score estimator against the jackknife estimates of the mean squared error of Lohr-Ybarra estimator of the Y_i's from the 30 small areas. The corrected score estimator performs only very slightly better than the Lohr-Ybarra estimator in this case. This could be attributed to the fact that the data used in this study has some departure from normality as is evident from Figure 3 in Chapter III. As we have commented earlier, the corrected score estimator is highly dependent on the normality assumption with respect to the data, which is violated in this case.

In Figure 5, we plot the jackknife estimates of the mean squared error for the Monte Carlo corrected score estimator against the jackknife estimates of the mean squared error of Lohr-Ybarra estimator of the Y_i's from the 30 small areas. The superiority of the Monte Carlo corrected score method is very evident here. The Monte Carlo method is highly robust to departures from normality as has been extensively exhibited in the simulation section of this chapter. Hence, as expected, the Monte Carlo method outperforms the Lohr-Ybarra method by a substantial margin in this study also.

Comparison of all the four methods in Figure 6 clearly shows the superiority of the Monte-Carlo based methods, namely SIMEX and MCCS, over the model-based ones.

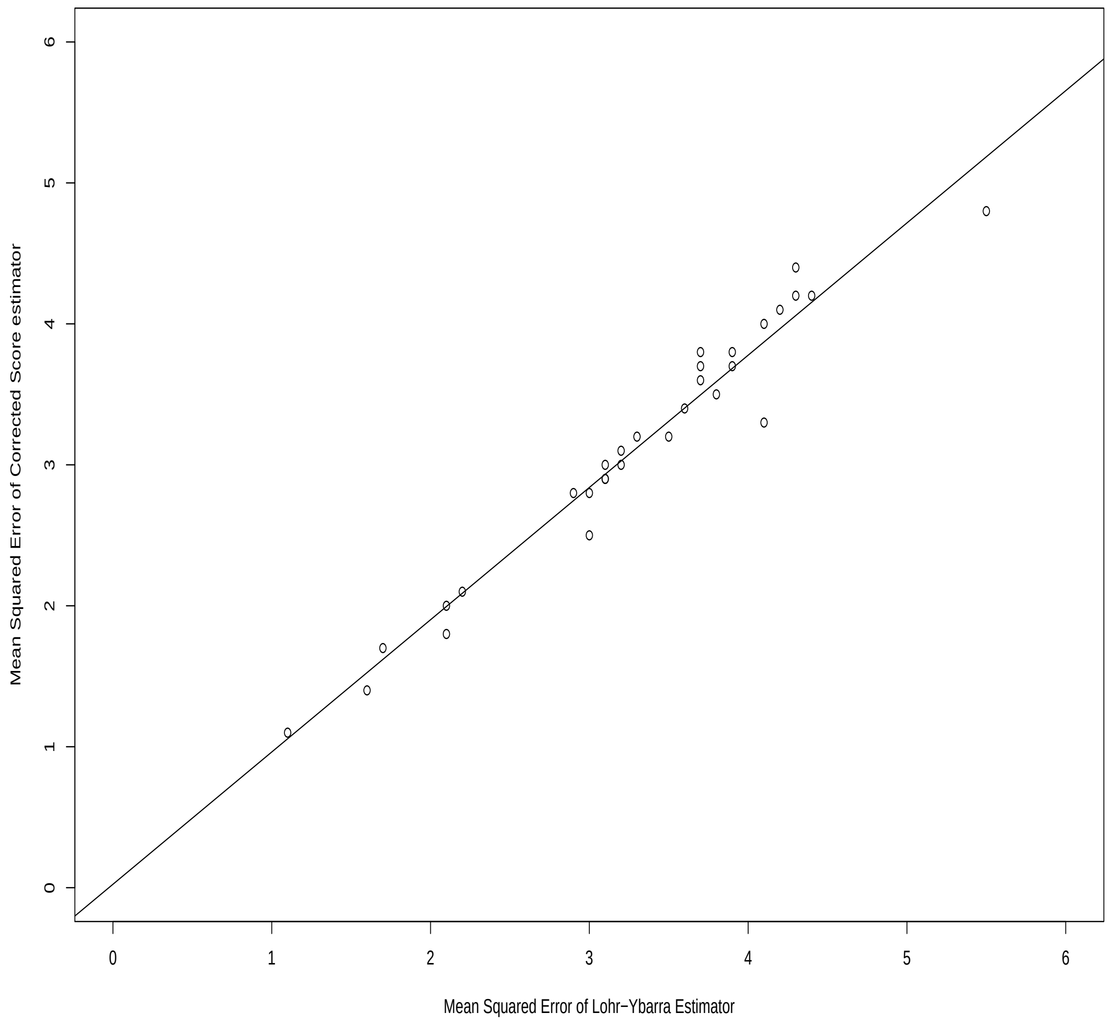

Fig. 4. Plot of jackknife estimates of mean squared error of the Lohr-Ybarra estimator ($\hat{Y}_{iME}$) of the mean body mass index, against the jackknife estimates of mean squared error of the corrected score estimator ($\hat{Y}_{iFHCS}$) for the 30 domains or small areas.

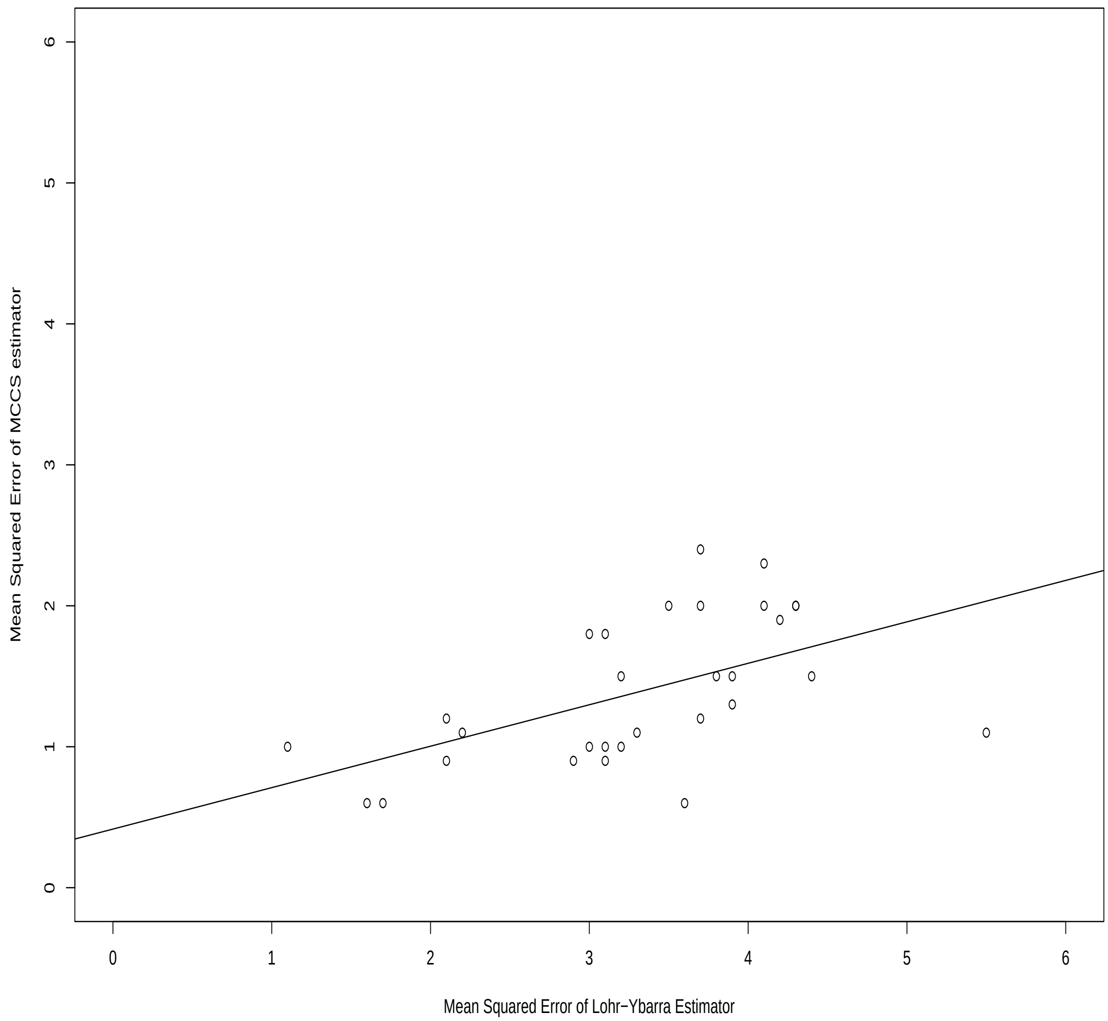

Fig. 5. Plot of jackknife estimates of mean squared error of the Lohr-Ybarra estimator ($\hat{Y}_{iME}$) of the mean body mass index, against the jackknife estimates of mean squared error of the Monte Carlo corrected score estimator ($\hat{Y}_{iMCCS}$) for the 30 domains or small areas.

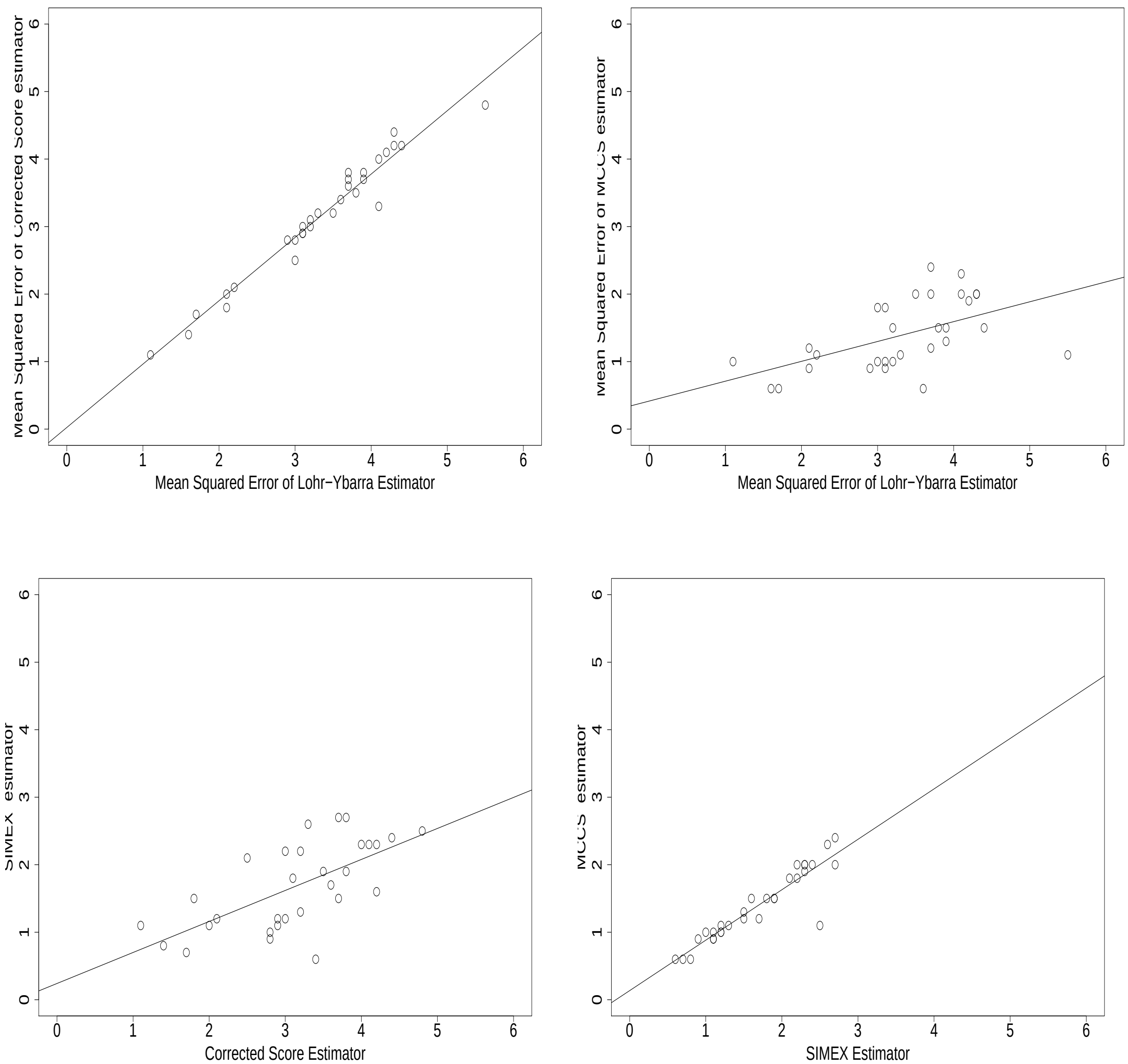

Fig. 6. Comparison of mean squared error for the four estimators. On the top left are the mean squared errors of the corrected score estimator versus those of the Lohr-Ybarra estimators. Top right: MCCS versus Lohr-Ybarra estimator. Bottom left: SIMEX versus corrected score estimator. Bottom right: MCCS versus SIMEX estimator.

CHAPTER V

SUMMARY AND CONCLUSIONS

As has been pointed out earlier, the literature with respect to dealing with measurement error in covariates for small area estimation is very limited. It is a topic that has started interesting survey statisticians only in the past few years. Lohr and Ybarra made the first step in this direction, but their estimator, though consistent, is very inefficient and does not really provide a solution to the problem of measurement error in covariates, beyond the fact that it was incorporated into the model. Torabi *et al.* (2009) also incorporated measurement error into the small area estimation model but they considered a unit level model instead and even though they provided a second error approximation of the mean squared error, their paper does not actually directly address the problem of measurement error.

This dissertation has directly dealt with the triple whammy of measurement error described in the introduction and the havoc it can wreak. There has been extensive research going on in the area of measurement error for decades now which has resulted in a wide variety of methods that answer the need for a solution to the measurement error problem. We make advantage of the flexibility and widespread applicability of these methods at our disposal and use them to tackle the problem with respect to small area estimation. Hence this dissertation is a confluence of the most appropriate methods available in the small area estimation literature as well as the literature associated with measurement error models.

Since our model is a structural measurement error model, that is, the covariates are treated as fixed constants, one of the most widely used methods in this scenario is simulation extrapolation (SIMEX) which we apply with great success to the Fay Herriot model with measurement error in covariates. This proves to be an effective method of bias correction in this case and provides reduction in the estimated mean squared errors compared to estimators that are currently available. The beauty of the SIMEX method lies in its effectiveness even when there is departure from normality or additivity of the measurement error. This is evident from the simulation results provided in Chapter II when data is generated from heavy-tailed t or skew-normal distributions. This gain in efficiency is also visible when SIMEX is applied to a data set consisting of body mass index values for certain demographic groups obtained from the 2003-2004 National Center for Health Statistics, National Health and Nutrition Examination Survey (NHANES), treating body mass index values from the corresponding National Health Interview Survey (NHIS) as auxiliary information. There is a strong case for applying measurement error bias correction methods in this case since the height and weight values used in the calculation of the body mass index values in the NHIS is reported by the respondents and not recorded by the interviewers. Application of SIMEX leads to a great reduction in the estimated mean squared errors of the small area estimates in this case. One fact stands out very clearly from this study, namely, when auxiliary information is available, a survey statistician should always make use of it to obtain small area estimates and not merely use the direct design-based estimators.

The second most widely used method of bias correction in structural measurement error models is that of corrected scores which we apply to the small area estimation model containing measurement error. We impose the normality assumption on the sampling errors, model errors and measurement errors and obtain corrected score estimates for the model. However, our model in this is more general than the Fay-Herriot model since it allows the random effect variances to be different from one small area to the next. The corrected score estimators are also provided for the Fay-Herriot model. In addition, we apply a simulation based technique of obtaining corrected scores called Monte Carlo corrected scores, which consists of generating pseudo-variables and adding them to the original covariates in a fashion similar, though not exactly the same as SIMEX. In out simulation studies we show that the simple corrected score estimator behaves well as long as the normality assumptions are not violated. But the Monte Carlo corrected score estimators and SIMEX estimators are able to handle departures from normality much better than the simple corrected score estimators. This behavior is evident when we obtain corrected score estimates and Monte Carlo corrected score estimates of the body mass index values in the data set mentioned earlier and compare the estimated mean square errors, since there is strong evidence of skewness in the data. It is also interesting to note that Lohr and Ybarra (2008) do not provide a method of estimating the measurement error variances in practical applications and treat it as known. We use data from the previous NHIS survey and treat it as a replicate of the error-prone covariate and then estimate the measurement error covariate as suggested by Carroll *et al.* (2006).

It is important to note that though SIMEX and Monte Carlo corrected scores are both simulation based methods, they are easy to implement. There is a SIMEX package available in R developed by Wolfgang and Lederer that is very convenient and flexible since it allows for different choices of extrapolant functions namely linear, quadratic or non linear. The Monte Carlo corrected score approach is straightforward to implement in any programming language that allows a complex-valued arithmetic.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1984), "On the Meaning of Within-factor Correlated Measurement Errors," *Journal of Data Science*, 3, 572–580.

Arellano-Valle, R. B., Bolfarine, H., & Lachos V. H. (2005), "Skew-normal Linear Mixed Models," *The Journal of Consumer Research*, 11, 415–438.

Armstrong, B. (1985), "Measurement Error in the Generalised Linear Model," *Communications in Statistics - Simulation and Computation*, 14, 529–544.

Battese, G. E., Harter, R.M. & Fuller, W. A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, 83, 28–36.

Bogue, D. J., & Duncan, B. D. (1959), "A Composite Method of Estimating Post Censal Population of Small Areas by Age, Sex and Color," *Vital Statistics- Special Report*, 47, No.6.

Buzas, J., & Stefanski, L. A. (1996), "Instrumental Variable Estimation in Generalized Linear Measurement Error Models," *Journal of the American Statistical Association*, 91, 999-1006.

Buzas, J., Stefanski, L. A., & Tosteson T. (2005), "Measurement Error," *Handbook of Epidemiology*, vol. 1, London: Springer.

Carroll, R. J., & Stefanski, L. A. (1987a), "Conditional Scores and Optimal Scores for Generalized Linear Measurement-error Models," *Biometrika*, 74, 703–716.

Carroll, R. J., & Stefanski, L. A. (1987b), "Covariate Measurement Error in Logistic Regression," *The Annals of Statistics*, 13, 1335–1351.

Carroll, R. J., & Stefanski, L. A. (1990), "Approximate Quasi-Likelihood Estimation in Models with Surrogate Predictors," *Journal of the American Statistical Association*, 85, 652–663.

Carroll, R. J., & Stefanski, L. A. (1997), "Asymptotic Theory for the Simex Estimator in Measurement Error Models," *Advances in Statistical Decision Theory and Applications*, 151–164.

Carroll, R. J., Kuchenhoff, H., Lombard, F. & Stefanski, L. A. (1996), "Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models," *Journal of the American Statistical Association*, 91, 242–250.

Carroll. R. J., Roeder. K., & Wasserman. L. (1999), "Flexible Parametric Measurement Error Models," *Biometrics*, 55, 44–54.

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models* (1st edition), London: CRC Press.

Carroll, R. J., Ruppert, D., Stefanski, L. A. , & Crainiceanu, C. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd edition), New York: Chapman & Hall.

Casella, G. & Berger, R. L. (2002), *Statistical Inference,* (2nd edition), Belmont, California: Thomson Learning.

Cheng, C. L., & Van Ness, J. W. (1999), *Statistical Regression with Measurement Error* (1st edition), London: Arnold.

Cook, J. R., & Stefanski, L. A. (1994), "Simulation Extrapolation in Parametric Measurement Error Models," *Journal of the American Statistical Association*, 89, 1314–1328.

Cook, J. R., & Stefanski, L. A. (1995), "Simulation Extrapolation: The Measurement Error Jackknife," *Journal of the American Statistical Association*, 90, 1247–1256.

Cox, D. R. & Hinkley, D. V. (1974), *Theoretical Statistics*, (1st edition), London: Chapman Hall.

Datta, G. S., & Lahiri, P. (2000), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems," *Statistica Sinica*, 10, 613–627.

Datta, G. S., Rao, J. N. K., & Smith, D. D. (2005), "On Measuring the Variability of Small Area Estimators Under a Basic Area Level Model," *Biometrika*, 92, 183–196.

Devanarayan, V. (1996), "Empirical Simulation Extrapolation Method for Heteroscedastic Measurement Error Models with Replicate Measurements," Ph.D. dissertation, North Carolina State University, Raleigh, NC.

Devanarayan, V. & Stefanski, L. (2002), "Empirical Simulation Extrapolation for Measurement Error Models with Replicate Measurements ," *Statistics & Probability Letters*, 59, 219–225.

Dick, P. (1995), "Modeling Net Undercoverage in the 1991 Canadian Census," *Survey Methodology*, 21, 45–54.

Eckert, R. S., Carroll, R. J. & Wang N. (1997), "Transformations to Additivity in Measurement Error Models," *Biometrics*, 53, 262–272.

Fay, R., & Herriot. R. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, 74, 269–277.

Fuller, W. A. (1980), "Properties of Some Estimators for the Errors-in-variables Model," *The Annals of Statistics*, 8, 407–422.

Fuller, W. A. (1987), *Measurement Error Models*, (1st edition), New York: Wiley.

Gimenez, P. & Bolfarine, H. (1997), "Corrected Score Functions in Classical Error-in-variables and Incidental Parameter Models," *Australian Journal of Statistics*, 39, 325–344.

Gimenez, P., Bolfarine, H., & Colosimo, E. A. (1987), "Hypotheses Testing for Error-in-variables Models," *Annals of the Institute of Mathematical Statistics*, 52, 698–711.

Gray, H. L., Watkins, T. A.; Schucany, W. R. (1973), "On the Jackknife Statistic and its Relation to UMVU Estimators in the Normal Case," *Communications in Statistics - Theory and Methods*, 2, 235–320.

Harville, D. A. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72, 320–340.

Henderson, C. R. (1963), "Selection Index and Expected Genetic Advance,", *Statistical Genetics & Plant Breeding, National Research Council,* 892, 141–63, Washington, D.C., National Academy of Sciences.

Henderson, C. R. (1975), "Best Linear Unbiased Estimation and Prediction Under a Selection Model," *Biometrics*, 31, 423–447.

Hossain, M. S. (2007), "Dealing with Measurement Error in Covariates with Special Reference to Logistic Regression Model: A Flexible Parametric Approach," Ph.D. dissertation, University of British Columbia, Vancouver, British Columbia.

Hwang, J. T. (1986), "Multiplicative Errors-in-variables Models with Applications to Recent Data Released by the U.S. Department of Energy," *Journal of the American Statistical Association*, 81, 680–688.

Iturria, S. J., Carroll, R. J. & Firth, D. (1999), "Polynomial Regression and Estimating Functions in the Presence of Multiplicative Measurement Error," *Journal of the Royal Statistical Society, Series B*, 61, 547–561.

Jiang, J., Lahiri, P., & Wan, S. (2002), "A Unified Jackknife Theory for Empirical Best Prediction with M-Estimation," *The Annals of Statistics*, 30, 1782–1810.

Jones, D. Y., Schatzkin, A., Green, S. B., Block, G., Brinton, L. A., Ziegler, R. G., Hoover, R., & Taylor, P. R. (1987), "Dietary fat and breast cancer in the National Health and Nutrition Survey I: Epidemiologic Follow-up Study," *Journal of the National Cancer Institute*, 79, 465–471.

Lechner, S. (2007), "The Multiplicative Simulation-extrapolation Approach," Konstanz, Germany, University of Konstanz (unpublished manuscript).

Lee, Y., & Nelder, N. A. (1996), "Hierarchical Generalized Linear Models (with discussion)," *Journal of the Royal Statistical Society, Series B*, 58, 619–678.

Lin, A. (1989), "Estimation of Multiplicative Measurement Error Models and Some Simulation Results," *Economics Letters*, 31, 13–20.

Lohr, S., & Ybarra, L. (2008), "Small Area Estimation When Auxiliary Information is Measured With Error," *Biometrika,* 95, 919–931.

Lyon, J. L., Alder, S. C., Stone, B. M., Scholl, A., Reading, J. C., Holubkov, R., Sheng, X., White Jr., G. L., Hegmann, K. T., Anspaugh, L., Hoffman, F. O., Simon, S. L., Thomas, B., Carroll, R. J., Meikle, A. W. (2006), "Thyroid Disease Associated With Exposure to the Nevada Nuclear Weapons Test Site Radiation: A Reevaluation Based on Corrected Dosimetry and Examination Data," *Epidemiology*, 17, 604–614.

MacMahon, S., Peto, R., Collins, R., Godwin, J., Cutler, J. , Sorlie, P., Abbott, R., Neaton, J., Dyer, A. & Stamler, J. (1990), "Blood Pressure, Stroke, and Coronary Heart Disease: Part 1, Prolonged Differences in Blood Pressure: Prospective Observational Studies Corrected for the Regression Dilution Bias," *Lancet*, 335, 765–774.

McCulloch, C. E., & Neuhaus, J. M. (2001), "Generalized Linear Mixed Models," *Encyclopedia of Biostatistics*, New York: Wiley.

Nakamura, T. (1990), "Corrected Score Function for Errors-in-variables Models: Methodology and Application to Generalized Linear Models," *Biometrika*, 77, 127–137.

National Center for Health Statistics, "1980 U.S. National Natality Survey Public Use Data Release," Atlanta, GA, National Center for Health Statistics, 1980.

National Center for Health Statistics, "2004 National Health Interview Survey (NHIS) Public Use Data Release: NHIS Survey Description," Atlanta, GA, National Center for Health Statistics, 2005.

National Center for Health Statistics, "Analysing and Reporting Guidelines: The National Health and Nutrition Examination Survey (NHANES)," Hyattsville, MD, National Center for Health Statistics, 2006.

Novick S., & Stefanski, L. A. (2002), "Corrected Score Estimation via Complex Variable Simulation Extrapolation," *Journal of the American Statistical Association*, 97, 472–481.

Prasad, N. G. N., & Rao, J. N. K. (1990), "The Estimation of Mean Squared Error of Small Area Estimators," *Journal of the American Statistical Association*, 85, 163–171.

Prentice, R. L. (1982), "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model," *Biometrika*, 69, 331–342.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, (1st edition), Wiley: New Jersey.

Rao, J. N. K. (2003a), "Some New Developments in Small Area Estimation," *Journal of the Iranian Statistical Society,*, 2, 145–169.

Rao, J. N. K. (2003b), *Small Area Estimation*, (1st edition), Hoboken, NJ: John Wiley Sons Inc.

Rivest, L. P., & Vandal, N. (1999), "Mean Square Estimations for Small Areas When the Small Area Variances Are Estimated," *Journal of Official Statistics*, 15, 1–24.

Robinson, G. K. (1991), "That BLUP Is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–32.

Schall, R. (1991), "Estimation in Generalized Linear Models with Random Effects,"*Biometrika*, 78, 719–727.

Scholtz, F. W. (1981), "Maximum Likelihood Estimation,"*Encyclopedia of Statistics*, (1st edition), New York: Wiley.

Shifa, N. (2009), "Estimation of QVF Measurement Error Models Using Empirical Likelihood Method," Ph.D. dissertation, Bowling Green State University, Bowling Green, OH.

Stefanski, L. A. (1985), "Unbiased Estimation of a Non Linear Function of a Normal Mean with Application to Measurement Error Models," *Biometrika*, 72, 583–592.

Stefanski, L. A. (1989), "Unbiased Estimation of a Non Linear Function of a Normal Mean with Application to Measurement Error Models," *Communications in Statistics*, 18, 4335–4358.

Stefanski, L. A., & Carroll, R. J. (1990), "Score Tests in Generalized Linear Measurement Error Models," *Journal of the Royal Statistical Society, Series B*, 52, 345–359.

Torabi, M., Datta, G. S., & Rao, J. N. K. (2009), "Empirical Bayes Estimation of Small Area Means Under a Nested Error Linear Regression Model with Measurement Error in the Covariates," *Scandinavian Journal of Statistics*, 36, 355–368.

Whittemore, A. S., & Keller, G. B. (1988), "Approximations for Regression with Covariate Measurement Error," *Journal of the American Statistical Association*, 83, 1057–1066.

Wolter, K. M. (2007), *Introduction to Variance Estimation*, (2nd edition), New York: Springer.

Wolter, K. M. & Fuller, W. (1982), "Estimation of Nonlinear Errors-in-variables Models," *The Annals of Statistics*, 10, 539–548.

Yi, G. (2000), "Covariate Measurement Error in Life History Data," Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Canada.

Zhong, X., Fung, W. & Wei, B. (2002), "Estimation in Linear Models with Random Effects and Errors-in-variables," *Annals of the Institute of Mathematical Statistics*, 54, 595–606.

APPENDIX A

A.1: Proof of Result 3.3.1

This result is from Lohr and Ybarra (2008). Define $\widetilde{Y}_i = a_i y_i + (1-a_i)\widehat{X}_i^T\beta$. Then, keeping in mind that $Y_i = X_i^T\beta + v_i$, $y_i = \widehat{X}_i^T\beta + r_i + e_i$, and $r_i = v_i + \left(X_i - \widehat{X}_i\right)^T\beta$ we have,

$$\begin{aligned}
MSE(Y_i - \widetilde{Y}_i) &= MSE\{X_i^T\beta + v_i - a_i(\widehat{X}_i^T\beta + r_i + e_i) - (1-a_i)\widehat{X}_i^T\beta\} \\
&= MSE(v_i - a_i r_i - a_i e_i) \\
&= MSE\{r_i - \left(X_i - \widehat{X}_i\right)^T\beta - a_i r_i - a_i e_i\} \\
&= MSE\{(1-a_i)r_i - \left(X_i - \widehat{X}_i\right)^T\beta - a_i e_i\} \\
&= (1-a_i)^2(\sigma_v^2 + \beta^T C_i\beta) + MSE(W) + {a_i}^2\psi_i,
\end{aligned}$$

where $W = \left(X_i - \widehat{X}_i\right)^T\beta$, is a constant in a_i. For minimization, taking partial derivative of $MSE(Y_i - \widetilde{Y}_i)$, equating to zero and solving for a_i we get:

$$a_i = \frac{\sigma_v^2 + \beta^T C_i\beta}{\sigma_v^2 + \beta^T C_i\beta + \psi_i}.$$

Now $y_i = \hat{X}_i^T\beta + r_i(\hat{X}_i^T, X_i) + e_i$ and $Y_i = X_i^T\beta + v_i$.

So the proof of $E(\tilde{Y}_{iME} - Y_i) = (1-\gamma_i)\{E(\hat{X}_i - X_i\}^T\beta$ is as follows:

$$\begin{aligned}
E(\tilde{Y}_{iME} - Y_i) &= E\{\gamma_i y_i + (1-\gamma_i)\hat{X}_i\beta - X_i^T\beta + v_i\} \\
&= \gamma_i E(y_i) + (1-\gamma_i)E(\hat{X}_i^T)\beta - X_i^T\beta \\
&= \gamma_i\{X_i^T\beta + (1-\gamma_i)E(\hat{X}_i^T)\beta - X_i^T\beta \\
&= (1-\gamma_i)\{E(\hat{X}_i - X_i\}^T\beta,
\end{aligned} \tag{A.1}$$

since $E(v_i) = 0$, $E(r_i) = (X_i - \hat{X}_i)^T\beta$ and $E(y_i) = X_i^T\beta$.

The proof of $MSE(\tilde{Y}_{iME}) = \gamma_i\psi_i$ follows from noting that $MSE(r_i) = \gamma_i\psi_i$.

A.2: Proof of Result 3.3.2

This description is done along the lines of Carroll *et al.* (1996). Let us first consider the *simulation step* of the SIMEX algorithm. Suppose one has an unknown parameter β that could be estimated using the estimating equation:

$$0 = \sum_{i=1}^{m} \phi(Y_{ib}, X_{ib}, \beta). \tag{A.2}$$

This estimating equation ϕ or method could be any that yields asymptotically normal estimators if we were to conduct the analysis as if measurement error did not exist, for example, least squares, weighted least squares, maximum likelihood, etc. In the simulations using the SIMEX R package, generalized least squares is used to compute estimates of β at each value of the pseudoerror ζ . Generalized least squares estimators are known to be consistent and asymptotically normal and we shall omit the details of those. A detailed discussion of their properties can be found extensively in statistical literature, like in McCulloch and Neuhaus (2001).

SIMEX operates as follows: Fix B (the number of iterations, then for each b^{th} iteration $(b = 1, ..., B)$ define $\widehat{\beta}_b(\zeta)$ as the solution to:

$$0 = \sum_{i=1}^{m} \phi(Y_{ib}, X_{ib}, \widetilde{X}_{ib}, \beta_b). \tag{A.3}$$

Now form the average of the $\widehat{\beta}_b(\zeta)$'s, namely:

$$\widehat{\beta_S}(\zeta) = B^{-1} \sum_{b=1}^{B} \widehat{\beta}_b(\zeta), \tag{A.4}$$

the subscript S pointing towards the simulation nature of the estimator. By standard estimation equation theory and under sufficient regularity conditions, $\widehat{\beta}_b(\zeta)$ converges in probability to $\beta(\zeta)$, where $\beta(\zeta)$ is the solution to:

$$\mathbf{0} = E\{\phi(\mathbf{Y}, \mathbf{X}, \widetilde{X}, \beta)\}. \tag{A.5}$$

For any fixed b, standard asymptotic theory yields the expansion:

$$m^{1/2}\{\widehat{\beta}_b(\zeta) - \beta(\zeta)\} = -\mathcal{A}^{-1}\{C_i, \zeta, \beta(\zeta)\} \times m^{-1/2} \sum_{i=1}^{m} \phi\{Y_{ib}, X_i, \widetilde{X}_i, \beta_b(\zeta)\} + o_p(1), \tag{A.6}$$

where $\mathcal{A}^{-1}\{C_i, \zeta, \beta(\zeta)\} = E\{(\frac{\partial}{\partial \beta})\phi(Y_i, X_i, \widetilde{X}_i, \beta(\zeta)\}$. This is because if $\widehat{\beta}_b(\zeta)$ is consistent, then by Taylor series approximation we have:

$$0 \approx m^{-1} \sum_{i=1}^{m} \phi\{\widetilde{Y}_i, \beta(\zeta)\} + \{m^{-1} \sum_{i=1}^{m} \frac{\partial}{\partial \beta} \phi\{Y_i, \beta(\zeta)\}$$

where $\beta(\zeta)$ is now the true parameter value. Applying the law of large numbers to the second term in the above equation inside the curly brackets we have:

$$\widehat{\beta}_b(\zeta) - \beta(\zeta) \approx -\mathcal{A}_m^{-1}(\beta) m^{-1} \sum_{i=1}^{m} \phi\{\widetilde{Y}_i, \beta(\zeta)\} \tag{A.7}$$

where $\mathcal{A}_m(\beta)$ is given by (A.9). Let $\mathcal{A}_m^{-T}(\beta) = \{\mathcal{A}_m^{-1}(\beta)\}^T$. Then $\widehat{\beta}_b(\zeta)$ is asymptotically normal with mean $\beta(\zeta)$ and covariance matrix $m^{-1}\mathcal{A}_m^{-1}\{\beta(\zeta)\}B_m\{\beta(\zeta)\}\mathcal{A}_m^{-T}\{\beta(\zeta)\}$ where

$$B_m\{\beta(\zeta)\} = m^{-1} \sum_{i=1}^{m} \mathrm{Cov}[\phi_i\{\widetilde{Y}_i, \beta(\zeta)\}] \tag{A.8}$$

and

$$\mathcal{A}_m\{\beta(\zeta)\} = m^{-1}\sum_{i=1}^{m} E\left\{\frac{\partial}{\partial\beta}\phi(\widetilde{Y}_i, \beta(\zeta)\right\} \tag{A.9}$$

Now the remainder term in (6) is of order $O(m^{-1/2+a})$ almost surely for any $a>0$. Because B is fixed the expression in (6) is uniform in $b = 1, ..., B$ and $\zeta \in \Omega$.

Let us define:

$$\chi_{B,i}\{C_i, \zeta, \beta(\zeta)\} = B^{-1}\sum_{b=1}^{B}\phi(Y_{ib}, X_{ib}, \widetilde{X}_{ib}, \beta_b). \tag{A.10}$$

Let $\mathcal{A}^{-1}(\cdot) = \mathcal{A}^{-1}\{C_i, \zeta, \beta(\zeta)\}$, then by Equation (A.6) we have

$$\begin{aligned} m^{1/2}\{\widehat{\beta}_b(\zeta) - \beta(\zeta)\} &= -\mathcal{A}^{-1}(\cdot)m^{-1/2}\sum_{i=1}^{m} B^{-1}\sum_{b=1}^{B}\phi\{Y_{ib}, X_{ib}, \widetilde{X}_{ib}, \beta_b(\zeta)\} \\ &= -\mathcal{A}^{-1}(\cdot)m^{-1/2}\sum_{i=1}^{m}\chi_{B,i}\{C_i, \zeta, \beta(\zeta)\} + o_p(1). \end{aligned} \tag{A.11}$$

The terms $\chi_{B,i}(\cdot)$ are mutually independent and identically distributed with mean zero (Carroll *et al.*, 1996) so that Equation (11) is simply a standard asymptotic linearization result for $\widehat{\beta}_b(\zeta)$.

Now we consider the step that consists of *combining the estimators* obtained at each value of the pseudoerror ζ. The SIMEX algorithm computes $\widehat{\beta}_b(\zeta)$ on a grid of values $\Omega = \{\zeta_1, ..., \zeta_M\}$. Denote the resulting vector of estimators by vectors $\{\widehat{\beta}(\zeta), \zeta \in \Omega\}$ by $\widehat{\beta}_{S^*}(\zeta)$ and corresponding vector of estimands by $\beta_*(\zeta)$. Define $\Psi_{B,i(1)}\{C_i, \Omega, \beta_*(\zeta)\} = vec[\chi_{B,i}\{C_i, \zeta, \beta(\zeta))\}]$, where $\zeta \in \Omega$ and $\mathcal{A}_{11}\{C_i, \Omega, \beta_*(\zeta)\} = diag[\mathcal{A}\{C_i, \Omega, \beta_*(\zeta)\}, \zeta \in \Omega]$. Then, using (11), the joint limit distribution of $m^{1/2}\{\widehat{\beta}_{S^*}(\zeta) - \beta(\zeta)\}$ is multivariate normal $(0, \Sigma)$ with,

$$\Sigma = \mathcal{A}^{-1}{}_{11}(\cdot)Q_{11}\{C_i, \zeta, \beta_*(\zeta)\}\mathcal{A}^{-T}{}_{11}(\cdot), \tag{A.12}$$

where

$$Q_{11}\{C_i, \zeta, \beta_*(\zeta)\} = \text{cov}\{\Psi_{B,1(1)}\{C_i, \Omega, \beta_*(\Omega)\}. \tag{A.13}$$

Now we consider the extrapolation step of the algorithm.

Having computed the vector of estimates $\widehat{\beta}_{S^*}(\Omega)$ on the grid $\Omega = \{\zeta_1, ..., \zeta_M\}$ (to understand the behavior of estimators for different amounts of measurement error) we now set $\beta(\zeta) = \{\mathcal{G}(\Gamma, \zeta), \zeta \in \Omega\}$ to the elements of $\widehat{\beta}_{S^*}(\Omega)$.

Define

$$\mathcal{G}'(\Gamma, \zeta) = \frac{\partial}{\partial \Gamma} \mathcal{G}(\Gamma, \zeta), \tag{A.14}$$

$$s(\Gamma) = \{\mathcal{G}'(\Gamma, \zeta_1), \mathcal{G}'(\Gamma, \zeta_2), ..., \mathcal{G}'(\Gamma, \zeta_M)\}, \tag{A.15}$$

and

$$R(\Gamma) = \widehat{\beta}_{S^*}(\Omega) - \{\mathcal{G}'(\Gamma, \zeta_1), \mathcal{G}'(\Gamma, \zeta_2), ..., \mathcal{G}'(\Gamma, \zeta_M)\}, \tag{A.16}$$

where $\mathcal{G}(\Gamma, \zeta)$ is a parametric model we fit to the $\beta_b(\zeta)$'s as a function of the ζ's. Various parametric models, such a linear, quadratic or non-linear, are possible with $\Gamma = (\alpha_0, \alpha_1, \alpha_2)^T$ and the non-linear model being $\mathcal{G}(\zeta, \Gamma) = \alpha_0 + \alpha_1(\lambda + \alpha_2)^{-1}$.

For any symmetric, positive-definite matrix F of weights and estimate $\widehat{\Gamma}$ of Γ is obtained by minimizing $R^T(\Gamma)F^{-1}R(\Gamma)$ which has estimating equation $0 = s(\Gamma)F^{-1}R(\Gamma)$. Taking F to be the identity matrix and and using the following three models (a) $\alpha_0 + \alpha_1\zeta(linear)$, (b) $\alpha_0 + \alpha_1\zeta + \alpha_2\zeta^2$ (quadratic) and (c) $\alpha_1(\zeta + \alpha_2)^{-1}$ (non-linear). Define $K(\Gamma) = s(\Gamma)F^{-1}s^{-1}(\Gamma)$ then,

$$\widehat{\Gamma} - \Gamma \approx Normal\{0, \Sigma(\Gamma)\}, \tag{A.17}$$

where $\Sigma(\Gamma) = K^{-1}(\Gamma)s(\Gamma)F^{-1}\Sigma F^{-1}s^T(\Gamma)K^{-1}(\Gamma)$.

Now the SIMEX estimator is,

$$\widehat{\beta}_{SIMEX} = \mathcal{G}(\widehat{\Gamma}, -1), \tag{A.18}$$

which, from the delta method, is asymptotically normal and has asymptotic variance matrix $\{\mathcal{G}'(\widehat{\Gamma}, -1)\}^T \Sigma(\Gamma)\{\mathcal{G}'(\widehat{\Gamma}, -1)\}$.

The fact that $\hat{\beta}_{SIMEX} \approx \beta + O_p(m^{-1/2})$ can be easily seen from the above results. But for a simpler discussion of this let us consider the simple linear model $Y = \beta_0 + \beta_x X + \epsilon$ with additive measurement error $\hat{X} = X + U$, U being independent of (Y, X) and has a mean zero and variance C. The ordinary least squares estimate of β_x, denoted by $\hat{\beta}_{x,naive}$, consistently estimates $\beta_x \sigma_x^2/(\sigma_x^2 + C)$ (Carroll *et al.* 2006) and thus is biased for β_x when $C > 0$.

Thus, for the SIMEX estimation method, the least squares estimate for the b^{th} data set , $\beta_b(\zeta)$ consistently estimates $\beta_x \sigma_x^2/\{\sigma_x^2 + (1+\zeta)C\}$ (Fuller, 1987). Buzas *et al.* (2005) studied asymptotically the bias of the nave least squares estimator and showed that $\hat{\beta}_x(\zeta) = \beta_x \sigma_x^2/\{\sigma_x^2 + (1+\zeta)C\} + o_p(m^{-1/2})$. Since the SIMEX estimator is obtained by extrapolating to the case of $\zeta = -1$, it follows that $\hat{\beta}_x(-1) = \beta_x \sigma_x^2/\{\sigma_x^2 + [1 + (-1)]C\} + o_p(m^{-1/2}) = \beta_x + O_p(m^{-1/2})$. This result can be easily extended to the Fay Herriot model since the only difference is the addition of random effect v_i from i^{th} small area to the simple linear model.

A.3: Proof of Result 3.3.3

Let $\beta \in \Theta$ denote the parameter of interest and let $\widehat{\beta}_{TRUE} = \mathbf{F}\{\{Y_i, \widehat{X}_i\}$, where $\widehat{X}_i$ is the covariate measured with error and $\mathbf{F}$ is a functional that maps the data set into Θ. Let $\widehat{\beta}_{NAIVE} = \mathbf{F}\{Y_i, \widehat{X}_i\}$. For $\zeta \geq 0$, define pseuso-variables as described in Section 3.2,

$$\widetilde{X}_{b,i} = \widehat{X}_i + \sqrt{\zeta} U_{b,i}.$$

Now define

$$\widehat{\beta}_b(\zeta) = \mathbf{F}\{\{Y_i, \widetilde{X}_i\},$$

and

$$\widehat{\beta}(\zeta) = E\{\widehat{\beta}_b(\zeta)|Y_i, \widetilde{X}_i\}.$$

It follows that $E\{\widehat{\beta}(\zeta)\} = E\{\widehat{\beta}_b(\zeta)\}$. By assumption, $\widehat{\beta}_b(\zeta)$ converges in probability to its expectation which we now treat as a function of the true estimand β_0. The variance of total measurement error in $\widetilde{X}_i$ is $(1+\zeta)C_i$. Denote this function by $\mathbf{F}\{\beta_0, (1+\zeta)C_i\}$. So by assumption, both $\widehat{\beta}_b(\zeta)$ and $\widehat{\beta}(\zeta)$ converge in probability to $\mathbf{F}\{\beta_0, (1+\zeta)C_i\}$.

Now if $C_i = 0$, (that is, no measurement error is present), then $\widehat{\beta}_{TRUE} = \widehat{\beta}_{NAIVE} = \widehat{\beta}_b(\zeta) = \beta(\zeta)$ (almost surely). So if $\widehat{\beta}_{TRUE}$ is a consistent estimator of β_0, then this implies that $\beta_0 = \mathbf{F}\{\beta_0, 0\}$.

Now, provided that $\mathbf{F}(\cdot, \cdot)$ is a continuous function, we can conclude that,

$$\begin{aligned}
\lim_{\zeta \to\ -1} E\{\widehat{\beta}(\zeta)\} &= \lim_{\zeta \to\ -1} \mathbf{F}\{\beta_0, (1+\zeta)C_i\} \\
&= \mathbf{F}\{\beta_0, 0\} \\
&= \beta_0.
\end{aligned}$$

The extrapolant function fit to $\widehat{\beta}(\zeta)$ for $\zeta > 0$ is approximating $E\{\widehat{\beta}(\zeta)\}$ and thus its extrapolation to $\zeta = -1$ (which is nothing but the SIMEX estimator) is approximating β_0. Hence $\widehat{\beta}_{SIMEX}$ is an approximately consistent estimator. An estimator is said to be approximately consistent if it converges in probability to some constant that is approximately equal to estimand.

Proof of Result 3.3.4:

The consistency of the estimator $\hat{\sigma}_v^2$ suggested follows from the fact that we have already shown that $\hat{\beta}_{SIMEX}$ is a consistent estimator and also from this well-known result: If $\{\hat{\theta}_1, ..., \hat{\theta}_n\}$ is a set of consistent estimators then $g(\hat{\theta}_1, ..., \hat{\theta}_n)$ is also a consistent estimator, provided g is a continuous function (Casella and Berger, 2002). It is easy to see that $\hat{\sigma}_v^2$ is a continuous function of $\hat{\beta}_{SIMEX}$, ψ_i, C_i and the observed data, so it will be consistent.

Proof of Result 3.3.5:

Along the lines of Lohr and Ybarra (2008), first let us note that,

$$\begin{aligned} E(\hat{X}_i y_i) &= E\{\hat{X}_i(\hat{X}_i^T\beta + r_i + e_i)\} \\ &= E\{(\hat{X}_i\hat{X}_i^T)\beta + \hat{X}_i r_i + \hat{X}_i e_i\} \\ &= E(\hat{X}_i X_i^T)\beta + E(\hat{X}_i r_i) + E(\hat{X}_i e_i) \\ &= (X_i X_i^T)\beta + E(\hat{X}_i v_i) + E(\hat{X}_i\beta) - E(X_i^T\beta) \\ &= (X_i X_i^T)\beta. \end{aligned}$$

since $r_i = v_i + (X_i - \hat{X}_i)^T\beta$, $E(\hat{X}_i) = X_i$ and $E(\hat{X}_i v_i) = E(\hat{X}_i e_i) = 0$ since we have assumed independence of X_i, v_i and e_i.

Also, $MSE(\hat{X}_i) = Var(\hat{X}_i) + [Bias(\hat{X}_i)]^2$, and since $MSE(\hat{X}_i) = C_i$ then it is easy to see that,

$$E(\hat{X}_i\hat{X}_i^T - C_i) = X_i X_i^T. \tag{A.19}$$

Now we split the mean squared error of the SIMEX estimator $\hat{Y}_{iSIM}$, by adding and subtracting $\widetilde{Y}_{iSIMEX}$, into three terms as follows:

$$
\begin{aligned}
MSE(\widehat{Y}_{iSIMEX}) &= E(\widehat{Y}_{iSIM} - Y_i)^2 \\
&= E(\widehat{Y}_{iSIMEX} - \widetilde{Y}_{iSIMEX})^2 + E(\widetilde{Y}_{iSIMEX} - Y_i)^2 \\
&+ 2E(\widehat{Y}_{iSIMEX} - \widetilde{Y}_{iSIMEX})(\widetilde{Y}_{iSIMEX} - Y_i) \qquad \text{(A.20)} \\
&= MSE(\widetilde{Y}_{iSIMEX}) + P + Q \\
&= \gamma_i \psi_i + P + Q,
\end{aligned}
$$

where $P = E(\widehat{Y}_{iSIMEX} - \widetilde{Y}_{iSIMEX})^2$, $Q = 2E(\widehat{Y}_{iSIM} - \widetilde{Y}_{iSIMEX})(\widetilde{Y}_{iSIM} - Y_i)$ and $MSE(\widetilde{Y}_{iSIMEX}) = \gamma_i \psi_i$ from this result (1) of Chapter II.

Now we can see that:

$$
\begin{aligned}
\widehat{Y}_{iSIMEX} - \widetilde{Y}_{iSIMEX} &= \{\widehat{\gamma}_i y_i + (1 - \widehat{\gamma}_i)\widehat{X}_i^T \widehat{\beta}_{SIMEX}\} - \{\widehat{\gamma}_i y_i + (1 - \gamma_i)\widehat{X}_i^T \beta\} \\
&= (\widehat{\gamma}_i - \gamma_i)(y_i - X_i^T \beta) + (1 - \widehat{\gamma}_i) X_i^T (\widehat{\beta}_{SIMEX} - \beta).
\end{aligned} \qquad \text{(A.21)}
$$

Now:

$$
\begin{aligned}
\widehat{\gamma}_i - \gamma_i &= \frac{\hat{\sigma}_v^2 + \hat{\beta}^T C_i \hat{\beta}}{\hat{\sigma}_v^2 + \hat{\beta}^T C_i \hat{\beta} + \psi_i} + \frac{\sigma_v^2 + \beta^T C_i \hat{\beta}}{\sigma_v^2 + \beta^T C_i \hat{\beta} + \psi_i} \\
&= \frac{\psi_i(\hat{\sigma}_v^2 + \hat{\beta}^T C_i \hat{\beta} - \sigma_v^2 + \beta^T C_i \beta)}{(\hat{\sigma}_v^2 + \hat{\beta}^T C_i \hat{\beta} + \psi_i)(\sigma_v^2 + \beta^T C_i \beta + \psi_i)} \qquad \text{(A.22)} \\
&= \frac{\psi_i(\hat{\sigma}_v^2 + \hat{\beta}^T C_i \hat{\beta} - \sigma_v^2 + \beta^T C_i \beta)}{(\sigma_v^2 + \beta^T C_i \beta + \psi_i)^2} + O_p(m^{-1}).
\end{aligned}
$$

The last step in the above equation can be explained as follows. We have already shown that,

$$
(\hat{\beta}_{SIM} - \beta) = O_p(m^{-1/2}),
$$

and $\hat{\sigma}_v^2 = \sigma_v^2 + O_p(m^{-1})$. So then

$$
m^{1/2}(\hat{\beta}_{SIMEX} - \beta)^T C_i (\hat{\beta}_{SIMEX} - \beta) m^{1/2} = O_p(1).
$$

Let

$$A^* = (\hat{\sigma}_v^2 + \hat{\beta}_{SIMEX}^T C_i \widehat{\beta}_{SIMEX} + \psi_i) - (\sigma_v^2 + \beta^T C_i \beta + \psi_i)$$
$$= (\hat{\sigma}_v^2 - \sigma_v^2) + (\widehat{\beta}_{SIMEX}^T C_i \hat{\beta}_{SIMEX} - \beta^T C_i \beta) + k,$$

where k is a constant and $(\hat{\sigma}_v^2 - \sigma_v^2)$ is known to be of order $O_p(m^{-1})$. Also note that,

$$\hat{\beta}_{SIMEX}^T C_i \hat{\beta}_{SIMEX} - \beta^T C_i \beta = (\hat{\beta}_{SIMEX} - \beta)^T C_i (\hat{\beta}_{SIMEX} - \beta) + 2\hat{\beta}_{SIMEX}^T C_i \beta$$
$$= O_p(m^{-1}) + o_p(m^{-1/2})$$
$$= O_p(m^{-1}).$$

Hence $(\hat{\sigma}_v^2 + \hat{\beta}_{SIMEX}^T C_i \hat{\beta} + \psi_i) = (\sigma_v^2 + \beta^T C_i \beta + \psi_i) + O_p(m^{-1})$.

Also we must note that:

$$E(y_i - \hat{X}_i^T \beta)^2 = E(v_i)^2 + E\{(\hat{X}_i - X_i)^T \beta (\hat{X}_i - X_i)\} + E(e_i^2)$$
$$= \sigma_v^2 + \beta^T C_i \beta + \psi_i,$$

and

$$E\{\hat{X}_i(y_i - \hat{X}_i^T \beta)\} = X_i X_i^T \beta - \{(X_i X_i^T + C_i)\beta\}$$
$$= -C_i \beta,$$

Using the results from before, i.e., $E(\hat{X}_i y_i) = (X_i X_i^T)\beta$ and $E(\hat{X}_i \hat{X}_i^T - C_i) = X_i X_i^T$. Hence,

$$\begin{aligned}
P &= E(\hat{Y}_{iSIMEX} - \tilde{Y}_{iSIMEX})^2 \\
&= E\{(\widehat{\gamma}_i - \gamma_i)(y_i - X_i^T\beta) + (1-\widehat{\gamma}_i)X_i^T(\widehat{\beta}_{SIMEX} - \beta)\}^2 \\
&= E\{(\widehat{\gamma}_i - \gamma_i)^2(y_i - X_i^T\beta)^2\} \\
&+ trE\{(1-\widehat{\gamma}_i)^2(\widehat{\beta}_{SIMEX} - \beta)(\widehat{\beta}_{SIMEX} - \beta)^T E(\hat{X}_i\hat{X}_i^T)\} \\
&+ 2E\{(\hat{\gamma}_i - \gamma_i)(1-\hat{\gamma}_i)(\widehat{\beta}_{SIMEX} - \beta)^T\}\{\hat{X}_i(y_i - \hat{X}_i^T\beta)\} \\
&= \frac{\psi_i^2 E(\hat{\sigma}_v^2 + \widehat{\beta}_{SIMEX}^T C_i \hat{\beta}_{SIMEX} - \sigma_v^2 + \beta^T C_i \beta)^2}{(\sigma_v^2 + \beta^T C_i\beta + \psi_i)^3} \\
&+ (1-\gamma_i)^2 tr\{B_m(C_i + X_i X_i^T)\} \\
&+ 2E\{(\hat{\gamma}_i - \gamma_i)(1-\hat{\gamma}_i)(\widehat{\beta}_{SIMEX} - \beta)^T\}(C_i\beta) + o(m^{-1}),
\end{aligned} \tag{A.23}$$

using the fact that $E(y_i - \hat{X}_i^t\beta)^2 = \sigma_v^2 + \beta^t C_i\beta + \psi_i$, $E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t = B_m$ and $E\{\hat{X}_i(y_i - \hat{X}_i^t\beta)\} = -C_i\beta$. Now,

$$\begin{aligned}
Q &= E\{(\hat{Y}_{iSIMEX} - \tilde{Y}_{iSIMEX})(\tilde{Y}_{iSIMEX} - Y_i)\} \\
&= E\{[(\hat{\gamma}_i - \gamma_i)(y_i - \hat{X}_i^T\beta) + (1-\hat{\gamma}_i)\hat{X}_i^T(\hat{\beta} - \beta)][\tilde{Y}_i - Y_i]\} \\
&= E(\hat{\gamma}_i - \gamma_i)E\{(y_i - \hat{X}_i^T\beta)(\tilde{Y}_i - Y_i)\} \\
&+ E\{(1-\hat{\gamma}_i)(\hat{\beta} - \beta)E[\hat{X}_i(\tilde{Y}_i - Y_i)]\} \\
&= E\{(1-\hat{\gamma}_i)(\hat{\beta} - \beta)^T\}(1-\gamma_i)C_i\beta,
\end{aligned} \tag{A.24}$$

since $E\{(y_i - \hat{X}_i^T\beta)(\tilde{Y}_i - Y_i)\} = 0$ through expansion and $E[\hat{X}_i(\tilde{Y}_i - Y_i)] = (1-\gamma_i)C_i\beta$. Using the expressions derived for P and Q above, we get result (5).

APPENDIX B

B.1: Proof of Result 4.3.1.

Now by property of corrected scores we know that $E^*\{l^*(\beta;\mathbf{V},\ \mathbf{W},\ \mathbf{Y})\} = l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})$. So from the law of large numbers or Kolmogorov theorem 2 stated by C.R. Rao (1973, Pg. 115) we know that $m^{-1}l^*(\beta;\mathbf{V},\ \mathbf{W},\ \mathbf{Y})\} = l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})$ converges to zero almost surely.

Now since we have assumed that β is identifiable, applying the strong law of large numbers and Jensen's inequality, we can easily see that

$$m^{-1}\{l(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})\} - l(\beta_t;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})\} < 0,$$

with probability one as $m \to \infty$ for $\beta \neq \beta_t$. This is because of the result by C.R. Rao (1973, Pg. 364) that states: If $logf(\beta;\mathbf{V},\ \mathbf{X},\ \mathbf{Y})$ is differentiable in an interval including the true value, the maximum likelihood equation has a root with probability one as $m \to \infty$ which is consistent for β.

This result states above and along the lines of C.R. Rao (1973, Pg.364-365) it is obvious that there exists a consistent root of $S^*(\beta;\mathbf{V},\ \mathbf{W},\ \mathbf{Y})$ with probability one as $m \to \infty$.

B.2: Proof of Result 4.3.3.

Since our model is $\mathbf{Y} = \mathbf{X}\beta + \mathbf{IV} + \epsilon$ and we observe $\mathbf{W} = \mathbf{X} + \mathbf{U}$, so along the lines of Zhong *et al.* (2002) we have:

$$\begin{aligned} m^{-1}\{\mathbf{W}^T R^{-1}\mathbf{W} - \mathbf{X}^T R^{-1}\mathbf{X} - tr(R^{-1})\Lambda\} &= m^{-1}\{(\mathbf{X}+\mathbf{U})^T R^{-1}(\mathbf{X}+\mathbf{U}) \\ &\quad - \mathbf{X}^T R^{-1}\mathbf{X} - tr(R^{-1})\Lambda\} \\ &= m^{-1}\{\mathbf{X}^T R^{-1}\mathbf{U} + \mathbf{U}^T R^{-1}\mathbf{X} + \mathbf{P}\}, \end{aligned}$$

where $\mathbf{P} = \mathbf{U}^T R^{-1}\mathbf{U} - tr(R^{-1})\Lambda$.

Since $\mathbf{U} \sim N(\mathbf{0}, I_m \otimes \Lambda)$, we have:

$$m^{-1/2}(\mathbf{X}^T R^{-1}\mathbf{U}) \sim N(\mathbf{0}, m^{-1}\mathbf{X}^T R^{-2}\mathbf{X} \otimes \Lambda).$$

Since we have assumed that the limit of $m^{-1}\mathbf{X}^T R^{-2}\mathbf{X}$ exists as $m \to \infty$, we can conclude that $m^{-1}(\mathbf{X}^T R^{-1}\mathbf{U}) = O_p(m^{-1/2})$. In the same way it can be shown that $m^{-1}(\mathbf{U}^T R^{-1}\mathbf{X}) = O_p(m^{-1/2})$.

Now let us consider $\mathbf{P}$ and define the element of $\mathbf{P}$ at the (a, b) position as:

$$P_{ab} = \sum_{k=1}^{m}\sum_{l=1}^{m} \mathbf{U}_{ka} R^{kl} \mathbf{U}_{lb} - \sum_{k=1}^{m} R^{kk}\Lambda_{ab},$$

where $\mathbf{U} = (\mathbf{U}_{ka})$, $R^{-1} = (R^{kl})$, $\Lambda = (\Lambda_{kl})$.

We can easily see that $E\{\mathbf{U}_{ka}\mathbf{U}_{kb}\} = \Lambda_{ab}$ and $E(P_{ab}) = 0$. Furthermore, we have:

$$\begin{aligned} E(P_{ab}^2) &= \sum_{k,l}\sum_{p,q} E(\mathbf{U}_{ka}, \mathbf{U}_{kb}, \mathbf{U}_{pa}, \mathbf{U}_{qb}) R^{kl} R^{pq} - \Lambda_{ab}^2\{tr(R^{-1})\}^2 \\ &= (\Lambda_{aa} + \Lambda_{ab}^2)\{tr(R^{-1}\}. \end{aligned}$$

The above result is obtained by adding the expected value terms which are non-zero and have equal indeces, and adding expected value terms that have pairwise equal indeces. All other terms will obviously be zero. We have assumed the limit of $m^{-1}tr(R^{-2})$ exists as $m \to \infty$. So $E(m^{-1/2}P_{ab})^2 = O_p(1)$ which means that $E(m^{-1}\mathbf{P}) = O_p(m^{-1/2})$.

So since $m^{-1}\mathbf{X}^T R^{-1}\mathbf{U}$, $m^{-1}\mathbf{U}^T R^{-1}\mathbf{X}$ and $E(m^{-1}\mathbf{P})$ are all of order $O_p(m^{-1/2})$ then it follows that:

$$\begin{aligned} m^{-1}\{\mathbf{W}^T R^{-1}\mathbf{W} - \mathbf{X}^T R^{-1}\mathbf{X} - tr(R^{-1})\Lambda\} &= m^{-1}\{\mathbf{X}^T R^{-1}\mathbf{U} + \mathbf{U}^T R^{-1}\mathbf{X} + \mathbf{P}\} \\ &= O_p(m^{-1/2}) \end{aligned}$$

$$\mathbf{W}^T R^{-1}\mathbf{W} = \mathbf{X}^T R^{-1}\mathbf{X} + tr(R^{-1})\Lambda + O_p(m^{1/2}).$$

B.3: Proof of Result 4.3.4.

Now corrected scores estimator of β is given by:

$$\hat{\beta}_{FHCS} = \{\mathbf{W}^t R^{-1}\mathbf{W} - tr(R^{-1})\Lambda\}^{-1}\mathbf{W}^t R^{-1}\mathbf{Y}$$

Using Result 3 we can say:

$$\begin{aligned} \widehat{\beta}_{FHCS} &= \{\mathbf{W}^t R^{-1}\mathbf{W} - tr(R^{-1})\Lambda\}^{-1}\mathbf{W}^t R^{-1}\mathbf{Y} \\ &= \{\mathbf{X}^T R^{-1}\mathbf{X} + O_p(m^{1/2})\}^{-1}\mathbf{W}^t R^{-1}\mathbf{Y} \\ &= \{I_p + O_p(m^{1/2})\}^{-1}(m^{-1}\mathbf{X}^T R^{-1}\mathbf{X})^{-1} m^{-1}\mathbf{W}^t R^{-1}\mathbf{Y} \\ &= \{I_p + O_p(m^{1/2})\}(m^{-1}\mathbf{X}^T R^{-1}\mathbf{X})^{-1} m^{-1}\mathbf{W}^t R^{-1}\mathbf{Y}, \end{aligned}$$

Since from Taylor's series expansion we can obtain the result $\{I_p + O_p(m^{1/2})\}^{-1} = I_p + O_p(m^{1/2})$. So we have:

$$\sqrt{m}\widehat{\beta} = \{I_p + O_p(m^{1/2}\}(m^{-1}\mathbf{X}^T R^{-1}\mathbf{X})^{-1}m^{-1/2}(\mathbf{W}^t R^{-1}\mathbf{Y}).$$

Let us define $\xi = \mathbf{W}^t R^{-1}\mathbf{Y}/\sqrt{m}$ and obtain its asymptotic properties. Let $R^{-1/2} = \Gamma\Omega\Gamma^T$ denote the spectral decomposition of $R^{-1/2}$ where $\Gamma\Gamma^T = I_m$ and $\Omega = diag(\lambda_1^{-1/2}, . ., \lambda_1^{-1/2})$ and λ_i's are the eigenvalues of R. Then we have:

$$\xi = \frac{\mathbf{W}^t R^{-1}\mathbf{Y}}{\sqrt{m}} = \frac{\mathbf{W}^t \Gamma\Omega\Gamma^T R^{-1/2}\mathbf{Y}}{\sqrt{m}} = \frac{\tilde{\mathbf{W}}^t R^{-1}\tilde{\mathbf{Y}}}{\sqrt{m}},$$

where

$$\tilde{\mathbf{W}} = \Gamma^T W \sim Normal(\Gamma^T\mathbf{X}, I_m \otimes \Lambda),$$

$$\tilde{\mathbf{Y}} = \Gamma^T R^{-1/2}\mathbf{Y} \sim Normal\{\Gamma^T R^{-1/2}\mathbf{X}\beta, R^{-1}(\sigma_v^2\Sigma + G)\}.$$

The k^{th} element of ξ is given by:

$$\xi_k = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\tilde{\mathbf{W}}_{ik}\lambda_i^{-1/2}\tilde{\mathbf{Y}} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\alpha_i.$$

Since the α_i's are independent and limit of $Var(\xi_k)$ exists as $m \to \infty$. So, by central limit theorem, ξ_k is asymptotically normal and so is $\widehat{\beta}_{FHCS}$. Moreover, we have assumed that the limit of $m^{-1}\mathbf{X}R^{-1}\mathbf{X}$ exists, then let $Q = m^{-1}\mathbf{X}R^{-1}\mathbf{X}$ and so from above we have:

$$\begin{aligned}
\sqrt{m}\hat{\beta} &= \{I_p + O_p(m^{1/2})\}(m^{-1}\mathbf{X}^T R^{-1}\mathbf{X})^{-1}m^{-1/2}\mathbf{W}^t R^{-1}\mathbf{Y} + O_p(m^{-1/2}) \\
&= (m^{-1}\mathbf{X}^T R^{-1}\mathbf{X})^{-1}m^{-1/2}\mathbf{W}^t R^{-1}\mathbf{Y} + O_p(m^{-1/2}) \\
&= Q^{-1}\xi + O_p(m^{-1/2}).
\end{aligned}$$

But since $E(\mathbf{W}^t R^{-1}\mathbf{Y}) = \mathbf{X}^t R^{-1}\mathbf{X}\beta$, then we have,

$$E(\xi) = \frac{\mathbf{X}^t R^{-1}\mathbf{X}\beta}{\sqrt{m}} = Q\sqrt{m}\beta.$$

substituting $Q = m^{-1}\mathbf{X}^T R^{-1}\mathbf{X}$. Hence $\sqrt{m}(\hat{\beta} - \beta)$ is asymptotically with mean zero.

To find the asymptotic variance, let us write:

$$\begin{aligned}\sqrt{m}(\hat{\beta} - \beta) &= Q^{-1}\xi - Q^{-1}Q\sqrt{m}\beta + O_p(m^{-1/2})\\ &= Q^{-1}\{\xi - E(\xi)\} + O_p(m^{-1/2}).\end{aligned}$$

Hence, $avar(\sqrt{m}\hat{\beta}) = Q^{-1}Var(\xi)Q^{-1}$. The variance of ξ can be obtained as follows:

$$\begin{aligned}Var(\xi) &= E^{+}\{Var^{*}(\xi)\} + Var^{+}[E^{*}(\xi)]\\ &= m^{-1}E^{+}\{\mathbf{Y}R^{-2}\mathbf{Y}\}\Lambda + m^{-1}Var^{+}\{\mathbf{X}R^{-1}\mathbf{Y}\}\\ &= m^{-1}\{tr[R^{-2}(\sigma_v^2\Sigma + G)] + \beta^T\mathbf{X}R^{-2}\mathbf{X}\beta\}\Lambda + m^{-1}\mathbf{X}^T R^{-2}\mathbf{X}(\sigma_v^2\Sigma + G),\\ &= m^{-1}\{\mathbf{J} + m^{-1}\mathbf{X}^T R^{-2}\mathbf{X}(\sigma_v^2\Sigma + G)\},\end{aligned}$$

since $\{tr[R^{-2}(\sigma_v^2\Sigma + G)] + \beta^T\mathbf{X}R^{-2}\mathbf{X}\beta\}\Lambda = \mathbf{J}$, $Var^{*}(\xi) = \{\mathbf{Y}R^{-2}\mathbf{Y}\}\Lambda$ and $E^{*}(\xi) = \mathbf{X}R^{-1}\mathbf{Y}$ are the conditional variance and expectation respectively of ξ with respect to $\mathbf{W}$ given $\mathbf{V}$ and $\mathbf{Y}$ and E^{+} and Var^{+} are the unconditional expectation and variance with respect to $\mathbf{V}$ and $\mathbf{Y}$.

So it follows that since $Var(\xi) = m^{-1}\{\mathbf{J}+m^{-1}\mathbf{X}^T R^{-2}\mathbf{X}(\sigma_v^2\Sigma+G)\}$, $Q = m^{-1}\mathbf{X}^t R^{-1}\mathbf{X}$ and $avar(\sqrt{m}\beta) = Q^{-1}Var(\xi)Q^{-1}$ then putting all these together, we get the desired asymptotic variance of $\hat{\beta}_{FHCS}$.

B.4.: Proof of Result 4.3.5

Now for the corrected score estimator of $\mathbf{V}$, $\mathbf{V}_{FHCS} = (I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{W}\hat{\beta}_{FHCS})$ we can see that:

$$\begin{aligned}
(\mathbf{V}_{FHCS} - \mathbf{V}) &= (I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{W}\hat{\beta}_{FHCS}) - (I+\Sigma^{-1})^{-1}(\mathbf{Y}-\mathbf{W}\beta) \\
&= -(I+\Sigma^{-1})^{-1}\{-(\mathbf{Y}-\mathbf{W}\hat{\beta}_{FHCS}) + \mathbf{Y}-\mathbf{W}\beta\} \\
&= -(I+\Sigma^{-1})^{-1}\mathbf{W}\{\hat{\beta}_{FHCS}-\beta\} \\
&= -\{m^{-1}(I+\Sigma^{-1})\}^{-1}\{m^{-1}\mathbf{X} + O_p(m^{-1/2})\}(\hat{\beta}_{FHCS}-\beta) \\
&= -\mathbf{J}_1^{-1}\mathbf{J}_2(\hat{\beta}_{FHCS}-\beta),
\end{aligned}$$

where β is the true value of the parameter, $\mathbf{J}_1 = m^{-1}(I+\Sigma^{-1})$ and $\mathbf{J}_2 = m^{-1}\mathbf{X}$. We also employ the result $\mathbf{W} = \mathbf{X} + O_p(m^{-1/2})$ whose proof is similar to that of result III.3.3. Then, in a fashion similar to result III.3.4. we get the desired result.

It is obvious from the last equation that the asymptotic variance of $\mathbf{V}_{FHCS}$ will be given by:

$$avar(\mathbf{V}_{FHCS} - \mathbf{V}) = \mathbf{J}_1^{-1}\mathbf{J}_2\{avar(\hat{\beta}_{FHCS})\}\mathbf{J}_2^T\mathbf{J}_2^{-1}.$$

VITA

NAME

Trijya Singh

Department of Statistics,

Texas A&M University,

3143 TAMU,

College Station, TX 77843-3143.

c/o Dr. Suojin Wang & Dr. Raymond J. Carroll.

EDUCATION

2011, Ph.D., Statistics, Texas A&M University

2006, Master of Arts, Statistics, University of Lucknow

2003, Bachelor of Arts, University of Lucknow

RESEARCH INTERESTS

Small Area Estimation, Measurement Error Models, Survey Sampling, Longitudinal Data, Non-Parametrics, Spatial Statistics.

The typist for this dissertation was Trijya Singh.